博士学位論文

Geometrical Structures Embedded in High Dimensional Data Sets and Deep Learning :Analysis and Application to Dynamical Systems （高次元データセットに潜む幾何構造と深層学習：その解析と大自由度力学系への応用）

本武　陽一

東京大学大学院
総合文化研究科広域科学専攻

指導教官: 池上　高志 教授

2016年3月

概要

近年，情報技術の急速な発展に応じて，これまでにない大規模で高次元なデータの収集が可能になってきた．同時に，これらのデータをいかに分析するかが問題となっている．一方で，我々の高次元データそのものに対する知見はまだ十分にあるとは言えない．本研究の目的は，画像データセットや時系列データセットのような高次元データの集合が持つ幾何的構造を観察することを通して，データの性質についてのより精緻な情報を得ることである．

本研究では，高次元のデータセットが低次元多様体上に分布するという多様体仮説を採用し，さらにそのデータセットでトレーニングされたDeep Neural Networks（以下　DNN）によって，その観測が可能であるという仮定を設定し，それを検証した．その上で大規模画像データセットや大規模時系列データセットの観測を行った．

その結果DNNは，データセット多様体を抽出する機能をもち，そこに誤差が生じる場合はあるものの，データセットの観察装置としておおよそ使用可能であることが判明した．そこで次に，DNNを観測装置として用いて大規模画像データセットの観察を行い，そのデータセットが多様体構造を持つことや，データのラベルの意味論的な構造とデータの幾何的構造が関係するといった，データに対する新しい理解に繋がる結果を得た．

さらに，画像以外に多様体構造を持つ大規模データセットとして，鳥や魚等の群れの集団行動を再現するボイドモデルに着目した．特に，本研究で分析対象としたのは，これまでの研究であまり対象とされてこなかった，複数の違う種類の群れが相互作用するような大規模な系であり，そこには,「そもそも群れをどのように定義し抽出するべきか」という本質的な問いかけが生起する．本研究ではこの課題に対して,「群れ」を「時系列データ中にある多様体構造」と定義することでその抽出を行い，そこで抽出された「群れ」について個別に分析を行った結果,単一の群れには見られない数多くの新しい知見を得た．

謝辞

本研究を進めるにあたり，ご指導を頂いた指導教員の池上 高志教授に感謝致します．迷走しがちな私の研究活動を我慢強くサポートして頂きました．また，日常の議論を通じて多くの知識や示唆を頂いた池上研究室の皆様に感謝します．そして，これまで 8 年間の研究生活を支えて下さった全ての方に感謝します．こうして研究を続けられたのは， 周りの皆様の温かい助言と励ましのお陰です．

目 次

図 目 次

表 目 次

第1章　序論

1.1　研究の背景

ケプラーの法則の発見は，言うまでもなく，中世ヨーロッパにとって，そして現代につながる自然科学的価値観の発達にとって，まさに分水嶺であった [34]．なぜならば，ケプラーの法則は近代的な意味での最初の「自然法則」であったからである．すなわち，特定の現象を支配する普遍的関係についての正確で証明可能な陳述であり，数学的用語によって表現された陳述であったのである．一方で，この革命の背景には，彼の師であるチコ・ブラーエによる大規模で組織的な天体運動の観察があった．チコ自身は，多くの画期的な発見を行ったというわけではない．しかし，彼の唯一にして最大の発見である,「天文学には正確で継続された精度の高い観測データが不可欠である」という事実の発見は，ケプラーやそれに続く現代科学の基礎となった [34]．チコ以前の，宗教的価値観に基づいて形成された，円と周転円とで組み立てられる欧州天文学（天空の幾何学）は，あまり多くの観測データを必要としなかったし，精度の高いデータを要求することもなかった．なぜなら円の決定には，中心とその円周上の一点がわかればよく，あるいは，中心が未知の場合には円周上の 3 点がわかればよい，という簡単な事情があったからである．実際，コペルニクスは,「天球の回転について」の書の全体を通じて，彼自身の観測値をわずか 27 しか記録していない．残りは，ヒッパルコスやプトレマイオスやその他の人のデータに頼っていた．チコに至るまでは，これが一般的なやり方だった [34]．しかし，チコの緻密で継続的な観察によって発見された矛盾に基づいて，ケプラーの法則は発見されたのである．

現代において我々は，天体現象に限らず，あらゆる領域において，当時とは比較にならない量と種類，そして非常に次元の高いデータを手に入れつつある．我々はそのデータを元に，試行錯誤をしながら，予測や分類・分析等に取り組んでいる．しかしそこで行われていることの多くは，工学的応用を目的としたもので，高いパフォーマンスがでるパラメータやアルゴリズムの開発を目的としており，データそのものに十分注意が払われているとは言い難い．さらに，多くの分析技術はパラメトライズされた確率モデルへのデータの当てはめを行っている．モデルの複雑さは違うものの，そこにはチコ以前の天体研究と同様，見落とされている事

実がある可能性がある．

本研究の目的はチコ同様に，データを観察してその構造や性質についての知見を収集することである．そして将来本研究が，どこかの天才が情報科学の世界における新たな分水嶺を見つけるための，足がかりとなることを期待するものである．

1.1.1 データセットの構造

本研究で観測する構造とは，高次元空間にあるデータセットがもつ分布の幾何的構造である（図 1.1 右）．データセットがどのような幾何構造を持つのかという点に関して，次のような仮説が提唱されている．

- データセットは，そのデータの次元 d より十分に低い次元 d_M を持つ（微分可能）多様体周辺に埋め込まれれている（詳細は次章）．

これは所謂，機械学習における「多様体仮説」と呼ばれるものであり [12][43]，これに基づいて開発された多くの機械学習アルゴリズムが高いパフォーマンスを達成できることから [56][59][50][7][65][51][48]，仮説の妥当性が経験的に支持されている．

一方で，直接的にこの仮説の妥当性の検証に取り組んだ研究はこれまでにほとんどなく，仮説成立の成否についての答えはまだ得られていない．

多様体仮説以外にも，半教師あり・教師なし学習領域でアルゴリズム開発に関連してデータセットの構造についての多様な検討がなされてきた．一例としては，低密度分離仮説や半教師あり平滑性仮説，クラスタ仮説といった，データセットの分布がもつクラスタ性とその意味的なクラスの関係性についての一連の仮説がある [13]．本研究では，このようなデータセットの構造についての仮説の多くは，次章で解説する拡張された多様体仮説によって包含されると考え議論を進めていく．

1.1.2 データセットの幾何構造の観察手段

それでは，このようなデータセットの幾何構造はどのような手法で観測できるだろうか．

本研究では，データセットの幾何構造の観測装置として，近年注目される機械学習技術である Deep Neural Networks（以下，DNN）を用いる．DNN は，以下の仮定が満たされる場合，前節で説明したデータセットの多様体構造を観測する装置として使用可能となる．

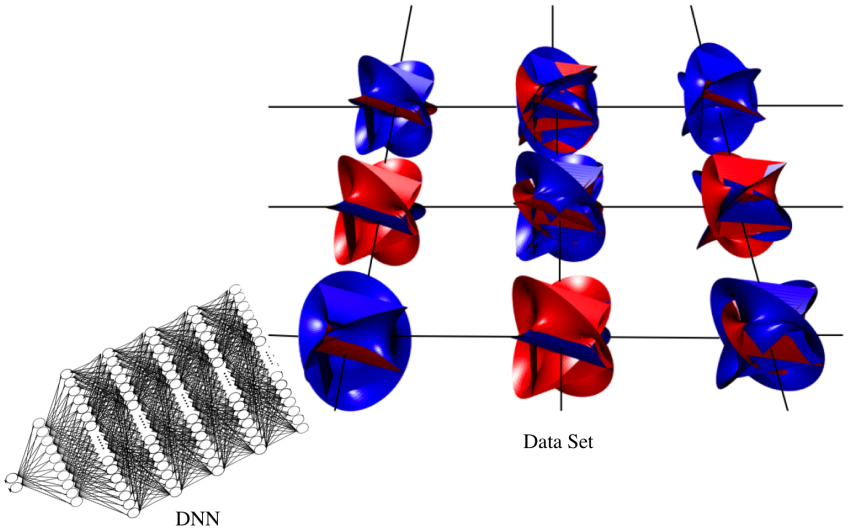

図 1.1: データセットの構造の観測のイメージ
天体を望遠鏡で観測するように，データセットの高次元空間中での分布構造（右）を DNN を用いて観測する（左）　.（図はカラビ・ヤウ多様体を 3 次元空間に写像したもの.）

- パフォーマンスの高い学習済み Deep Neural Networks は，上の多様体を多様体と同じ次元のユークリッド空間（図 1.2 参照．以降，”大域的な座標系 ”と呼称）へ写像する機能をもつ.

この仮説は，入江ら [71] や Hinton ら [30][15] によって部分的な検証が試みられている．しかしながら，この仮説の妥当であるかについての明確な答えはまだ得られていない.

この仮説が成り立つ場合に，具体的にどのようにしてデータセットの幾何構造についての情報を取得するかについては第 3 章で述べる.

この仮説は人工ニューラルネットに対して設定されたものであるが，脳神経科学者であるデカルロらは，人間の視覚経路に対して同様の仮説を提唱している [21]. この仮説では，IT 野においてこの大域的座標系が実現されているとしている．後述するように，この仮説における DNN や神経回路網の行う写像は単射（N 対 N 写像）であり，おばあちゃん細胞仮説 [27][6] のような N 対 1 写像によるデータの抽象化とは違った描像を与える.

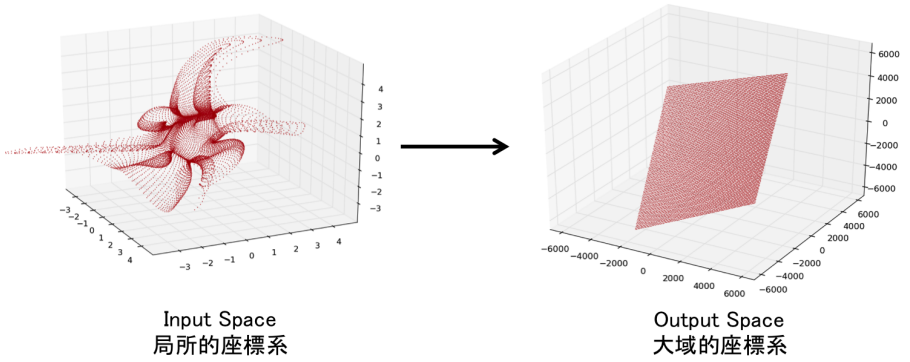

図 1.2: 大域的座標系
図の変換では，局所的な座標系の集合として表現される多様体（左図赤点）が，位置によらず，同じ座標系を共有する大域的な座標系へ変換されている（右図赤点）

1.2 本論文の概説

本論文ではまず，DNN がデータセットの幾何構造の観測装置として利用可能であるかを検証する．

そこでまず，高次元空間中の低次元多様体（n 次元球）上に分布する人工的なデータセットを作成し，出力層のノード数をその多様体次元と一致させてある DNN（具体的には，Deep Auto Encoder[30]）をトレーニングした上で，それを分析した．その結果，DNN によって多様体の次元や接空間の構造などがおおよそ観測可能であることが確認された．同時に，データセットの幾何構造の種類によっては，観測結果に誤差が生じることも確認された．

さらに，実践的データセットである手書き文字データセット MNIST[37] を用いて，さらに実際的な状況下での仮説の検証を行った．MNIST データセットは，先行研究によって多様体のおおよその次元などの構造が一部判明しており，それとの比較によって DNN の測定装置としての機能の妥当性の検証が行える．その結果，実践的なデータセットにおいても DNN が観測装置として利用可能であることを支持する結果が得られた．

これらの結果を踏まえ，大規模自然画像データセット ImageNet[20] を学習した DNN の分析を行い，データセットの構造の観測を試みた．この結果，ImageNet データセットが多様体仮説を満たすことと整合性のある結果が得られ，さらに，多様体の次元や曲がり具合といったデータセットの幾何的な構造が，画像の意味的なクラスタやその階層構造（「秋田犬」→「犬」といったカテゴリの階層）と関係することが示唆された．本研究ではこの結果を説明する仮説として，「画像の意味的な階層とデータセットの階層的幾何的構造が関係する」という仮説を設定し，そ

れを検証した．具体的には，幾何学的な階層構造を持ったデータセットを作成した上で，そのデータセットで学習されたDNNを解析した．その結果，DNNが多様体の階層的な幾何構造を折り紙のように段階的に展開するという，観測された現象と矛盾しない現象が生じている可能性が示唆された．

最後に，多様体仮説とは別の文脈でデータセットからの多様体構造の抽出を試みてきた，大自由度力学系の縮約構造の抽出という課題に対して，DNNを応用することを試みた．具体的には，鳥の群れのモデルであるReynoldsのボイドモデル[46]の大規模なシミュレーション結果に対してこれを適用した．そこには複数の群れが複雑に相互作用するような状況が存在し，これまでに行われてきたような方法での群れの抽出は難しい．そこで本研究では,「群れ」の定義を,「ある同一の多様体に，その運動の時系列データが埋め込まれた個体の集合」と仮定し，その多様体の抽出によって「群れ」の抽出をすることを試みた．その結果，違う性質を持った複数の「群れ」の同時抽出に成功した．

1.3 本研究の特色

本論文の特色は，以下の点となる．

1. 大規模高次元なデータセットをありのままに観察しようという目的を設定し，具体的にそれを試みた点．

2. DNNのデータセット観測装置としての性能検証のため，3次元以上の多様体構造をもつ人工データを用いて，DNNがデータセットの多様体構造を大域的な座標系へ写像する機能をもつことを定量的に確認した点．そして，観測結果に誤差が生じるケースについての検討も行った点．

3. 実践的なデータセット（MNIST）においても，DNNが多様体を大域的な座標系へ写像する機能をもつことを確認した点．

4. 階層的な意味構造をもつImageNetデータセットを観察し，それが多様体構造をもつことを示した点．および，意味的な階層とデータセットの幾何構造が関係することを示唆した点．

5. 上を説明する仮説として，データセットが階層的幾何構造をもち，それと意味的な階層構造が対応するという仮説を提案し，初歩的な検証を行った点．

6. 複数の群れが相互作用するような複雑な大自由度力学系において,「ある同一の多様体に，その運動の時系列データが埋め込まれた個体の集合」として

「群れ」を定義することで，性質の違う複数の群れを同時に抽出することに成功した点.

1.4 本論文の構成

本論文の構成は以下の通りである.

第 2 章では，多様体仮説と機械学習の関係性について概説した上で，DNN をデータセットの観察装置として用いる上での前提となる仮説と本研究の目的について説明する.

第 3 章では，まず DNN の観測装置としての性能の評価を人工データを用いて行う．その上で，実践的なデータセットでトレーニングされた DNN を解析し，そこで得られたデータセットの幾何構造についての考察とその評価を行う.

第 4 章では，DNN の大自由度力学系への応用について説明する．そのためにまず，具体的な応用対象である大自由度ボイドモデルについて解説し，DNN による分析結果との比較のために，既存手法による解析を行う．その上で，深層学習を大自由度ボイドモデルへ応用した結果を示す.

最後に第 5 章において総合考察を行う.

第2章　高次元データと多様体仮説

2.1　多様体仮説

この世界に存在するデータは，おおよそどのような性質を持つのであろうか？

ここに，機械学習の応用の中で実証されてきた 1 つの仮説がある，機械学習における多様体仮説 [12] である．以下，この仮説について解説する．

まず，多様体の数学的定義を確認する．多様体は，曲線や曲面といった，我々が慣れ親しんでいる図形の任意次元への一般化であり，以下で定義される．

Definition 2.1.1 M は以下の定義を満たす時，m 次元微分可能多様体という．[41]

1. M は位相空間である．
2. M には対の族 $\{(U_i, \phi_i)\}$ が与えられている．
3. $\{U_i\}$ は M を被覆する開集合写像である．すなわち $\bigcup_i U_i = M$．ϕ_i は U_i から $\mathbb{R}^m$ の開部分集合 U_i の上への同相写像である．
4. $U_i \cap U_j \neq \emptyset$ を満たす U_i と U_j が与えられたとき，$\psi_{ij} = \phi_i \circ \phi_j^{-1}$ は $\phi_j(U_i \bigcap U_j)$ から $\phi_i(U_i \bigcap U_j)$ への無限階微分可能な写像である．

正確さを欠いて要約すると，曲面上の接平面として定義されるユークリッド空間を，曲面全体にわたって稠密に貼りあわせてできる空間が多様体である．

4 を含まない単なる位相多様体ではなく，微分可能性を導入したのは，データセットの構造として，滑らかな多様体を想定していることと，後で用いる分析手法に必要な前提となる為である．

ここでは，図 2.1 のような空間でデータセットの分布を考える．データセットが画像であれば，空間の次元は画像のピクセルに対応し，空間中の 1 点が 1 つの画像に対応する．例えば，画像データが 10,000 ピクセルであれば，空間は 10,000 次元となる．この高次元空間に分布するデータセットが，空間の次元に比べて非常に低次元な多様体に埋め込まれているだろうという仮説が多様体仮説である．この仮説は，次のような考えに基づいている．

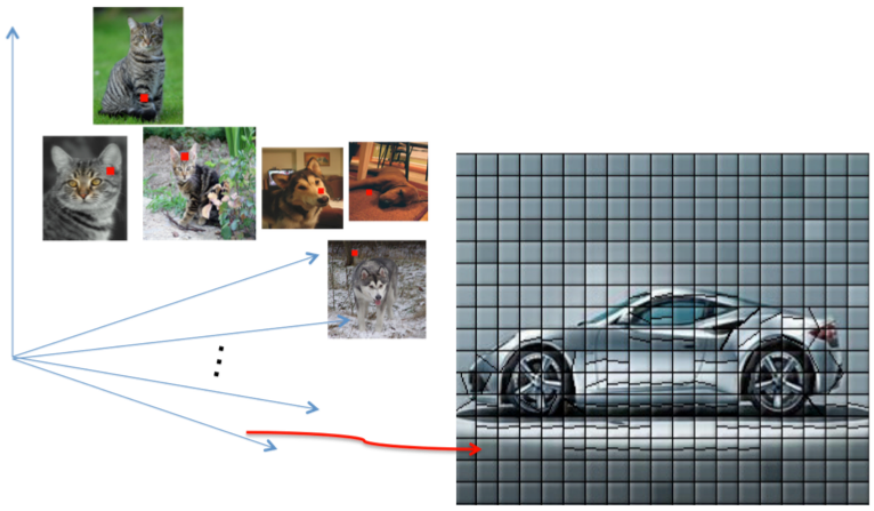

図 2.1: 多様体の存在する空間：データが画像であれば 1 つの次元を 1 つのピクセルに対応させる空間．この空間では 1 画像が 1 点で表される．

実世界に存在するデータは，視点の変化や物体自身の運動，手書き文字データ等であれば個人差などによって，少しずつ変化したものの集合となる．その変化が連続的であれば，図 2.2 のように，同一クラスのデータが曲線や曲面上に分布することになる．この曲面や曲線が多様体になると考えられるのである．

ここでは具体的に，Rifai[47] に基づき，多様体仮説を以下のように定義する．

Hypothesis 2.1.1 高次元空間に存在する実世界のデータは，非常に低次元の非線形多様体付近に集中している． ([12] [43]).

Hypothesis 2.1.2 高次元空間に存在する実世界のデータは，クラス（カテゴリ）毎に違う部分多様体に埋め込まれており，それらの部分多様体の間は低密度領域となっている．(分類問題に対する多様体仮説)

仮説 2.1.2 は，図 2.3 のように手書き数字データセット分布がそれぞれの数字毎に別々の連結空間に分離されることと説明できる．ちなみに，図 2.3 は手書き数字データセット（MNIST）を，後述する多様体学習アルゴリズム（t-SNE）を利用して 3 次元空間に圧縮した結果である，

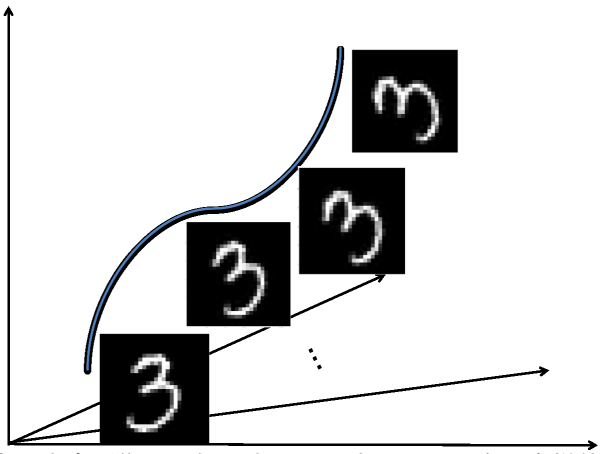

図 2.2: 手書き文字画像の回転に応じて形成される 1 次元多様体のイメージ

2.2 機械学習と多様体仮説

これまでに，多様体仮説に基づく機械学習アルゴリズムや手法が数多く開発され，成功をおさめている．このことは，多様体仮説の妥当性を支持する．そこで，以下にそれらの概要を解説する．

2.2.1 semi-supervised 学習

semi-supervised 学習と呼ばれる，ラベルあり・なしデータを混在させ学習し，精度を高めようとする一連の研究がある．これらの研究の中に，多様体仮説に基づいた手法がある．例えば，Bengio ら [13] は，多様体仮説を前提に，データの類似度に基づくグラフィカルモデルを構築し，ラベルを伝搬させることで学習データを増加させることを実現している．

2.2.2 多様体学習

多様体仮説を前提に，データセットの多様体構造を大域的座標系へ写像する機能を明示的に獲得しようというのが，多様体学習アルゴリズムである．複数のアルゴリズムが提案されているが，ここでは，本研究とも関係してくる ISOMAP（Isometric feature mapping）と t-SNE について概説する．

ISOMAP[56] は，多様体上で多次元尺度構成法（MDS）を適用し，次元圧縮を行う手法である．多様体上での距離の算出には，多様体の定義にのっとり，近傍

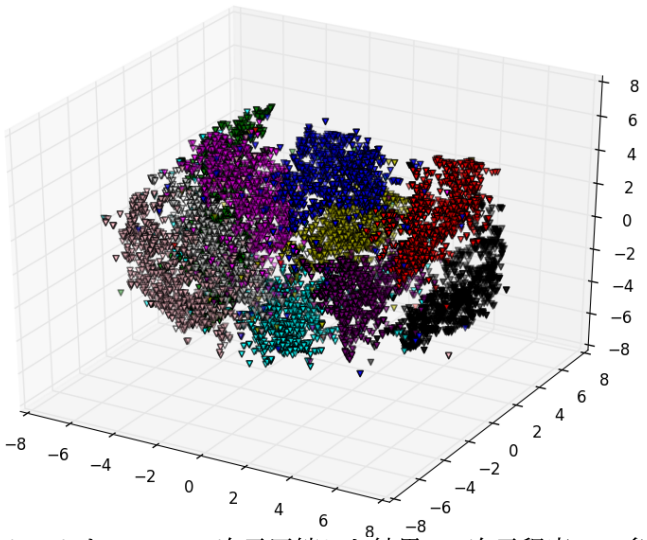

図 2.3: MNIST データセットを t-SNE で次元圧縮した結果．3 次元程度で，多様体構造を抽出できているようにみえる．

はユークリッド距離で近似し，それ以外は近傍の距離をもとに最短パス探索をするという手法を用いる．

t-SNE[59] は，多様体上での距離を確率分布として表現する手法である．圧縮前の確率分布には正規分布を，圧縮後の確率分布には t 分布を用いる．その上で，圧縮前後の確率分布の KL-dovergence を最小化するように分布のモデルパラメータを決定するのである．図 2.3 は，この手法による次元圧縮の結果である．

この他にも，Locally Linear Embedding（LLE）[50]，Laplacian Eigenmaps[7]，Semidefinite Embedding（SDE）[65] などといったアルゴリズムが提案されている．

2.2.3　ニューラルネットワーク：多層パーセプトロン

Simard ら [51] は，接線伝搬法という手法を適用することで，パーセプトロンの学習に多様体仮説の考え方を導入した．

次章で述べるように，ニューラルネットワークの写像関数のヤコビアンは，データ点まわりでの微分に対応する．今，データ点が多様体上に位置しているとすると，多様体と垂直方向の微分値は 0 にならなければならない．そこで，接線伝搬法では，ヤコビアンの各成分の和を正則化項として利用し，低次元多様体の抽出を試みている．

2.3 深層学習と多様体仮説

Hinton [29] の Deep Neural Networks の有効な学習法の発見以来，深層学習と呼ばれる，DNN を用いたパターン認識分野が脚光を集めている．例えば，Quoc らは，youtube からランダムに抽出した大量の画像を Deep Neural Networks（以下 DNN）に学習させることで,「猫の顔」といったカテゴリを自動で抽出することに成功した [36]．また，Szegedy らは，10 以上の層を持たせた DNN を用いることで，非常に高い画像認識の精度を達成している [54].

このような深層学習において，Rifai らは，多様体仮説に基づく深層学習アルゴリズムを開発した．それは Contractive-Auto-Encoder と呼ばれているアルゴリズムで，simard ら [51] の接線伝搬法と同様に，Auto-Encoder（付録 A.1 参照）の損失関数にヤコビアン各成分の正則化項を加えることで実現される [48]．このアルゴリズムによって，Rifai らは高いパフォーマンスを獲得しており，実際にヤコビアンの特異値分布（次章で解説）が，他のアルゴリズムと比較して急峻になり，その多くが 0 となることも示し DNN が多様体を捉えているとした.

このような Rifai の研究 [48] 等を背景にして，Bengio らは DNN による表象学習と多様体仮説の関係性に言及している [9].

2.4 本研究の目的

Contractive-Auto-Encoder のような，多様体仮説を前提にした深層学習アルゴリズム以外の方法で，DNN に多様体から大域的座標系への写像関数を学習させることはできないのであろうか.

入江 [71] らは，3 次元球上に分布するデータセットを作成し，入力と出力を一致させるような学習を，中間層のノード数が多様体の次元と一致する 2 であるような多層パーセプロトンで行った．そして，その中間層の 2 個のニューロンを格子状に刺激した際の，出力層での発火パターンが，学習データである球の上を滑らかに覆い尽くすことを確認した．そして，この結果からその DNN がデータセットの多様体構造を大域的な座標系へ写像する関数を獲得できると結論した.

また，hinton らの Deep-Auto-Encoder（付録 A.1 参照）を用いた研究でも，やはり中間層が 2 次元となるような DNN において，手書き文字画像や文章データなどが，クラス毎に分離した形でマッピングされることが示された．この事実は，データセットに多様体構造があり，それを DNN が大域的な座標系へ写像していることを示唆していると考えられる [30][15].

一方で以上の研究は，多様体の存在や，DNN がそれを写像する機能を持ち得ることを示唆するに留まっており，直接的・定量的にその存在を確認したものとは言

い難い．特に，2 次元以上の多様体については，これらの可視化に頼った手法で検証することが難しいと考えられる．一方，前節で述べた Rifai の Contractive-Auto-Encoder の研究では [48]，次章で述べるような定量的な方法によって，提案モデルによって多様体を捉えていることを示しているが，分析の対象は提案手法に限られており，かつ他手法と比較して多様体を捉える傾向があることを示しているだけで，直接的に多様体仮説を示すような分析は行われていない．そこで，本研究では，多様体仮説を満たすデータセットで学習された DNN が，多様体構造を大域的な座標系に写像する機能をもつことを，定量的・網羅的に調べることを第 1 の目的，実世界のデータセットが多様体構造をもつか，そしてそれは具体的にどのような構造を持つかを観察することを第 2 の目的とする．

また，DNN が多様体を大域的座標系へ展開する機能を持つことが確認される場合，それは多様体抽出機として利用出来る．DNN は前述した多様体学習アルゴリズムと比較して，写像関数自体を学習できるという利点がある．このことは，大規模なデータセットにある多様体構造の抽出が可能になることを意味している．従って，この利点を生かした応用について検討することを第 3 の目的とする．

第3章　観測装置としてのDeep Neural Networksの性能検証

3.1　DNN写像関数の分析による多様体構造の推定法

前章で見たように，高次元データセットは低次元多様体周辺に分布しており，パフォーマンスの高いDNN（Deep Neural Networks）は，それを多様体と同じ次元の大域的な座標系（定義は第1章参照）に写像する機能を持つことが示唆されている．このように，ニューラルネットワークが獲得した関数を，データセット多様体を大域的な座標系へ写像する関数だとみなすと，その関数を解析することで元の多様体の性質を知ることが可能となる．なぜならば，多様体から多様体への写像の微分は，以下で定義される多様体の接空間を定義し，そこから多様体の次元や接ベクトル等の情報を得ることができるからである．

ここで結論のみ簡単に説明すると，ニューラルネットワークの写像関数の微分（ヤコビアン行列）の特異値・特異ベクトル（非正方行列における固有値・固有ベクトルに相当．DNNは入力と出力の次元が違うため，ヤコビアンが非正方行列になる.）のうち，0より大きな特異値に対応する特異ベクトルが多様体の接線方向を，0の特異値に対応するベクトルが多様体の垂直方向をあらわす．従って，0でない特異値の数から，多様体の次元もわかる．また，特異ベクトルには，右と左があり，右特異ベクトルが入力空間で表現された多様体の水平・垂直ベクトルをあらわし，左特異ベクトルは出力空間で表現された多様体の水平・垂直ベクトルをあらわす．

以下，接（ベクトル）空間と写像関数の関係について詳しく説明していく．まず，接空間は，次のように説明される．

今，n次元ユークリッド空間$\mathbf{R}^n$に存在するm次元多様体をMとする．この多様体上のC^∞級曲線を考えると，この曲線上のある点pにおいて，接線を考えることができる．ここで，多様体M上にあって，$M \to \mathbf{R}^m$という写像を行う関数をfとする．これらを用いると，曲線に沿って運動する粒子の，$\mathbf{R}^m$空間での速度

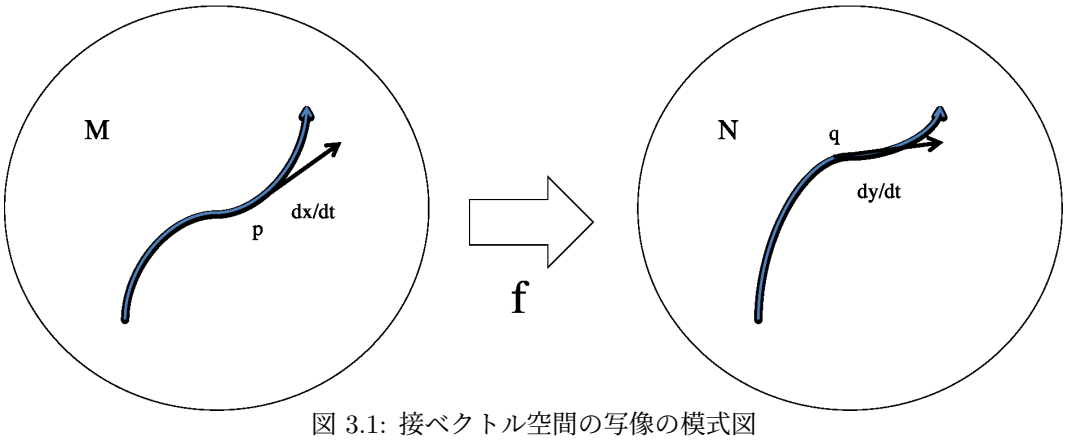

図 3.1: 接ベクトル空間の写像の模式図

ベクトルを定義でき，それは以下のようになる．

$$\vec{v} = \frac{dc}{dt}|_{t=0} = (\frac{dx_1}{dt}(0), \frac{dx_2}{dt}(0), \cdots, \frac{dx_m}{dt}(0)) \quad = \quad \frac{d(f \circ c)}{dt}(0)$$

ここで，$x_1, x_2, \cdots, x_m$ は，M 上のある点近傍の局所座標系（$\mathbf{R}^m$）で c をあらわしたものである．

ある点 p を通る全ての曲線について，この速度ベクトルを集めた集合を接ベクトル空間という．3 次元空間中の球（2 次元多様体）の例でいえば，球の接平面が，接ベクトル空間に対応する．

次に，接ベクトル空間から接ベクトル空間への写像を考える．M,N をそれぞれ m 次元，n 次元の C^r 級多様体，$f : M \to N$ を C^r 級写像とする（$1 \leq r \leq \infty$）．ここで，点 p を通る M 上の C^r 級曲線，

$$c : (-\epsilon, \epsilon) \to M, c(0) = p \tag{3.1}$$

を考える．

この曲線を写像 f でうつすと，図 3.1 のように，点 $q = f(p)$ を通る N 上の C^r 級曲線

$$f \circ c : (-\epsilon, \epsilon) \to N, f \circ c(0) = q \tag{3.2}$$

が得られる．この 2 つの曲線上の速度ベクトルの集合が，それぞれ接ベクトル空間を形成するので，これらの速度ベクトルの関係性から，写像 f による，2 つの接ベクトル空間の関係が判明する．

そこでまず，曲線 $f \circ c$ の速度ベクトルを求めることを考える．

ここで，点 p を含む M の座標近傍を $(U; x_1, \cdots, x_m)$，点 $q = f(p)$ を含む N の座標近傍を $(V; y_1, \cdots, y_m)$ とすると，この座標近傍での写像 f を局所座標表示で記述すると，

$$\begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \cdots, x_m) \\ \cdot \\ \cdot \\ f_n(x_1, x_2, \cdots, x_m) \end{pmatrix} \tag{3.3}$$

となる．従って，$f \circ c$ は，

$$f \circ c(t) = \begin{pmatrix} f_1(x_1(t), x_2(t), \cdots, x_m(t)) \\ \cdot \\ \cdot \\ f_n(x_1(t), x_2(t), \cdots, x_m(t)) \end{pmatrix} \tag{3.4}$$

で与えられる．これより，曲線 $f \circ c$ の速度ベクトルを求めると，

$$\begin{aligned} \begin{pmatrix} \frac{dy_1}{dt}|_{t=0} \\ \cdot \\ \cdot \\ \frac{dy_n}{dt}|_{t=0} \end{pmatrix} &= \begin{pmatrix} \frac{d}{dt} f_1(x_1(t), x_2(t), \cdots, x_m(t))|_{t=0} \\ \cdot \\ \cdot \\ \frac{d}{dt} f_n(x_1(t), x_2(t), \cdots, x_m(t))|_{t=0} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p)\frac{\partial x_1}{\partial t}(0) + \frac{\partial f_1}{\partial x_2}(p)\frac{\partial x_2}{\partial t}(0) + \cdots + \frac{\partial f_1}{\partial x_m}(p)\frac{\partial x_m}{\partial t}(0) \\ \cdot \\ \cdot \\ \frac{\partial f_n}{\partial x_1}(p)\frac{\partial x_1}{\partial t}(0) + \frac{\partial f_n}{\partial x_2}(p)\frac{\partial x_2}{\partial t}(0) + \cdots + \frac{\partial f_n}{\partial x_n}(p)\frac{\partial x_m}{\partial t}(0) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_m}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(p) & \cdots & \frac{\partial f_n}{\partial x_m}(p) \end{pmatrix} \begin{pmatrix} \frac{dx_1}{dt}|_{t=0} \\ \cdot \\ \cdot \\ \frac{dx_m}{dt}|_{t=0} \end{pmatrix} \end{aligned}$$

となる．これは，多様体 M の接空間から多様体 N 接空間への変換が，ヤコビアンで表現されることを意味する．

今具体的に考えているのは，M 次元ユークリッド空間中の m 次元部分多様体から，N 次元ユークリッド空間中の n 次元部分多様体への写像である．（ただし，インプット空間でデータが多様体近傍にノイズを持って分布していることを除いて，m 次元部分多様体と，n 次元部分多様体の次元は，同じ n であるとする．）

従って，ヤコビアンの rank は n となり，その固有ベクトルの集合が，多様体の接ベクトル空間を形成する．つまり，ヤコビアンの 0 でない固有値の個数が，多様体の次元を表す．ニューラルネットにおける具体的な演算は，次のようになる．単層のニューラルネットにおける写像関数は，例えば，活性化関数をシグモイド関数とすると，

$$y_j(t+1) = f(\sum_i (x_i(t) \cdot W_{ij}(t)) + B_j(t)) \tag{3.5}$$

$$f(x) = 1/(1+e^{-gx}) \ \ (g : const) \tag{3.6}$$

となる．従って，ヤコビアンは，以下で定義される．

$$\begin{aligned} J &= \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_1} & \cdots & \frac{\partial y_N}{\partial x_N} \end{pmatrix} \\ &= \begin{pmatrix} y_1(1-y_1)w_{11} & \cdots & y_1(1-y_1)w_{N1} \\ \vdots & \ddots & \vdots \\ y_1(1-y_1)w_{1N} & \cdots & y_N(1-y_N)w_{NN} \end{pmatrix} \end{aligned}$$

多層ニューラルネットワークの写像のヤコビアンは，微分の連鎖則から，単層ヤコビアンの積として，以下で導出される．例えば，L 層のニューラルネットの場合，

$$J_{all} = J(0) \cdot J(1) \cdots J(L-1) \tag{3.7}$$

となる．

このヤコビアンは，入力層と出力層のノード数が同数でない限り正方行列とはならない．従って，行列の分解には特異値分解（SVD：Singular Value Decomposition）を用いる．この分解にによって得られる特異値と特異ベクトルによって，多様体の接空間の情報を得ることができる．左特異ベクトルが入力空間での多様体の接線方向を，右特異ベクトルが出力空間での多様体の接線方向を与える．

データセットの多様体構造を大域的な座標系へ写像するためには 2 つの機能が必要となる．1 つは複雑に折りたたまれた多様体構造を大域的な座標系へマッピングするために，折りたたまれた多様体を押し広げる機能である．2 つ目は，M 次元ユークリッド空間の次元を圧縮して n 次元多様体を見つけ出す機能である．正確に 2 つの機能を切り分けて観察することはできないが，おおまかにはヤコビアン

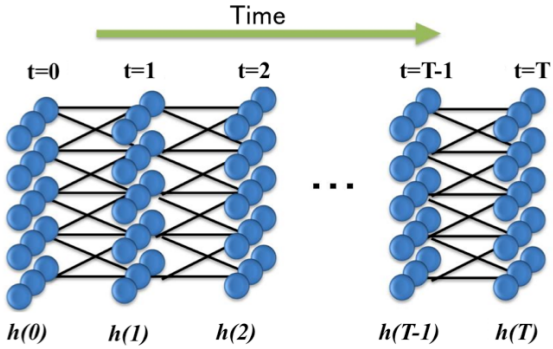

図 3.2: 多層ニューラルネットワークの時間発展

の左特異ベクトルによって多様体を押し広げる機能を観察でき，右特異ベクトルによって次元を圧縮する機能をみることができると考えられる．

後の機能の理解に有用な為，ここで上記の演算のもう一つの見方を提示しておく．それは，力学的な視点である．DNN の時間発展として 2 つのものが考えられる．1 つは，学習中の重みの時間発展である．もう 1 つは，図 3.2 のように DNN の各階層を時間に対応付け，層が進むに従って変化するニューロンの発火パターンの時間発展を考える視点である．ここでは後者の視点でニューラルネットワークをみる．

従って，ニューロン発火の時間発展は，次式で定義される．

$$h_j(t+1) = f(\sum_i (h_i(t) \cdot W_{ij}(t)) + B_j(t)) \tag{3.8}$$

$f(x)$ としては，よくシグモイド関数，

$$f(x) = 1/(1+e^{-gx}) \quad (g : const) \tag{3.9}$$

が使われる．ここで，$h_i(t)$ は t 層の隠れ層のノード状態を，$W_{ij}(t)$ は，t 層から $t+1$ 層の間の重み行列を，$B_j(t)$ は第 t 層のバイアス値を表すものとする（図 1 参照）．また，i,j は各層のノードのインデックスになっている．

この時間発展方程式は，図 2.1 のような画像の 1 ピクセルを 1 次元とする空間上で，画像に対応する粒子が，層を発展するとともにどのように移動するかを表現している．

この時間発展に対して，第 t 層における粒子位置の摂動に対する，$t+1$ 層での

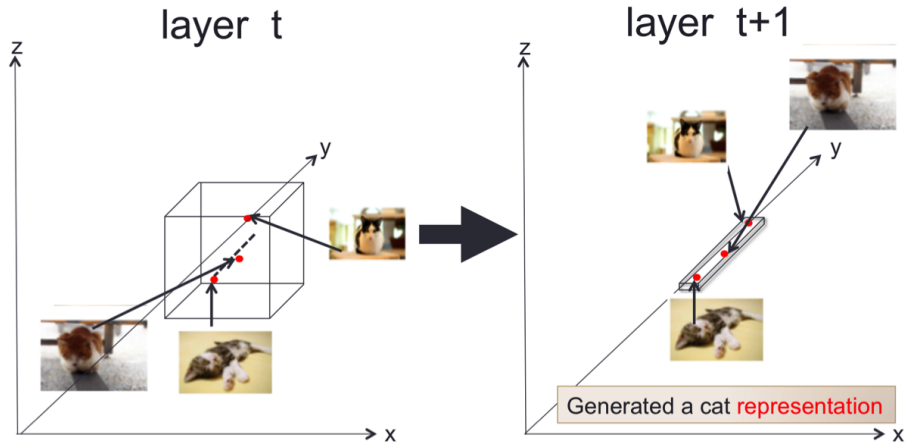

図 3.3: ニューラルネットワークのダイナミクスとカテゴリの形成

変動を表すヤコビアン行列が，以下で定義される．

$$J(t) = \begin{pmatrix} \frac{\partial h_1(t+1)}{\partial h_1(t)} & \cdots & \frac{\partial h_1(t+1)}{\partial h_N(t)} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_N(t+1)}{\partial h_1(t)} & \cdots & \frac{\partial h_N(t+1)}{\partial h_N(t)} \end{pmatrix} \tag{3.10}$$

このヤコビアンの特異値・特異ベクトルを求めることによって，どの方向（特異ベクトル）への摂動が保存され（特異値 > 1），どの方向が消去される（特異値 $<< 1$）かが分かる．図 3.3 のような場合を考えると，layer-t の 3 次元空間中に 3 つのネコ画像がのる 1 次元多様体（線）があり，それが layer-t+1 への時間発展で余分な 2 次元が圧縮され　ネコ多様体が抽出されるという表像である．

3.2　検証の方針

前章で設定した目的に基づき，ここまで説明してきた手法を用いて以下のような分析を行った．

まず，DNN が多様体構造を大域的な座標系へ写像する機能をもちうることを，人工的に生成したデータを用いて検証した．

次に，より実践的なデータである MNIST 手書き文字データセットによってトレーニングされた DNN を分析し，実践的な状況においても仮説が満たされるかを検証した．

3.3 DNN観測装置の性能検証1：人工データ

ここでは，DNNがデータセット観測装置として必要な条件を，どの程度満たすかを検証する．つまり繰り返しになるが，DNNが高次元（n次元）空間中にある$m+1$次元球の構造を持ったm次元多様体を，m次元の大域的な座標系に写像するような関数を獲得できるか検証する．具体的には，学習済みDNN写像関数のヤコビアンのrank（0でない特異値の数）によって，DNNが多様体の次元を検出できるかを検証し，ヤコビアンの右特異ベクトルによって，DNNが多様体の接線方向・垂直方向をとらえているかを検証した．出力空間での多様体の接空間をあらわす左特異ベクトルについては，図3.4のOutput Layer Spaceように，大域的座標系が実現されている場合，その垂直ベクトルが全て同じ方向を向く必要がある．しかし本研究では，中間層が多様体の次元と一致するDNNを用いるため，中間層が大域的座標系になっていることは自明である．従って，中間層を多様体の次元と一致させた検証実験では，左特異ベクトルの分析は行わなかった．

3.3.1 実験方法

本実験では，Deep Auto Encoderを用いた（付録A参照）．Deep Auto Encoderは，中間層のノード数が相対的に少ない砂時計型のネットワーク構造をしたDNNで，出力が入力を再現するように学習される．本実験においても，m次元多様体（m+1次元球）上に分布するデータセットを入力とし，それらを再現するように学習を行った．

学習の損失関数として，二乗誤差関数

$$E(\mathbf{W}) = \sum_{i=1}^{n} |x_i - y(x_i)|^2 \tag{3.11}$$

を用い，活性化関数としてシグモイド関数

$$f(x) = 1/(1+e^{-gx}) \ \ (g : const) \tag{3.12}$$

を用いた．

本実験では，図3.5にあるようなDeep Auto Encoder[10-20-10-2-10-20-10]を用いた．このネットワークを，n次元空間中の半球面（m次元多様体）に分布するデータセットを用いてトレーニングした．図3.6の左図は，3次元空間中の3次元球（2次元多様体）の場合のデータセット（サンプル数：1,000）の例である．本実験のためには,「n次元空間中に存在するm次元多様体（m+1次元球，$n > m$）」

の学習が必要である．そこでまず，半球上に分布するデータを $m+1$ 次元空間中の $m+1$ 次元半球（$m>2$）として作成した（サンプル数：1,000，正確に多様体上に分布するのではなくノイズを加えて多様体付近に分布させる）．さらに，空間の次元を $n=10$ と設定し，$m+1$ 次元空間に存在する球データを 10 次元空間中に分布するデータにする為に，10 次元の各軸まわりにランダムな角度 $\theta_i, (i=1,\cdot\cdot,10)$ で回転させ，10 次元空間に写像した．

このように作成した学習データセットを用いて，Deep Auto Encoder [10-20-10-m-10-20-10]（m:多様体の次元）をトレーニングした．また分析は，先ほど行った各軸まわりの変換の逆変換を行うことで元の $m+1$ 次元空間で行った．

以下，説明の便宜のため，Deep Auto Encoder[10-20-10-m-10-20-10] の各層を，[input - layer1 - layer2 - layer3 - layer4 - layer5 -layer6] と呼称する．

次に具体的な分析の方法を，図 3.6 の例を用いて説明していく．図 3.6 右図は，入江ら [71] の研究と同じ，3 次元空間中の 3 次元球データセットを用意し，同じノード数の Deep Auto Encoder をトレーニングした結果である．多様体の次元が 2 以下であるため，入江ら [71] の研究と同様にして，中間層のニューロンをグリッド状に刺激することで，中間層が学習した座標系を可視化できる．可視化の結果が図 3.6 中の赤点である．この結果より，DNN が半球多様体を大域的な座標系へ写像する関数を学習しているとわかる．また，中間層の空間が半球の頂点付近をおおよそ滑らかにマッピングされていることより，DNN が半球頂点付近を大域的な座標系に変換できていることがわかる．

この DNN のヤコビアンから右特異ベクトルを算出した．入力空間が 3 次元なので，右特異ベクトルは 3 つ算出される．その中で特異値が 0 となる特異ベクトルが，多様体の垂直方向成分である．この垂直ベクトルが，データセットの半球多様体の垂直ベクトルと一致することが確認できれば，ヤコビアンが多様体の接空間を捉えているということができる．そこで，図 3.7 左のような基準ベクトルとの偏角を計算した．基準ベクトルは，半球の底面に沿ったベクトルと定義し，データセットから算出される垂直ベクトルと基準ベクトルとの偏角を θ，右特異ベクトルと基準ベクトルとの偏角を φ とした．すると，図 3.7 右のように，半球頂点の θ=90 °付近で $\varphi=\theta$ となる分布が観察された．一方，半球の切断面付近（θ=0 °,180 °）に向かうにつれて，$\varphi=\theta$ からずれていった．この原因は中間層の空間の多様体のマッピングが図 3.8 のように一様でなかったためであると考えられる．つまり，半球の頂点付近に比べて半球の切断面付近では，空間が引き伸ばされた形になっていたため，多様体の接空間も変形し，$\varphi=\theta$ からずれてしまったと考えられる．入江ら [71] の研究でも半円多様体の切断面付近で予想からのずれが観測されている（入江らの研究ではこの誤差は無視している.）．ネットワーク構造やその他のハイパーパラメタを変えた実験を行った場合にも同様に切断面付近が引

き伸ばされる現象が観察された．この現象の原因は幾何学的なものである．つまり，正方形となる 0 ∼ 1 の連続値をとる 2 次元の中間層で，半球のような構造を被覆しようとする場合，正方形を伸縮させる必要が生じるのである．この結果として，前述の現象が生じたものと考えられる．

以上より，半球頂点の θ=90 °付近では，DNN が多様体を大域的な座標系へ写像することが確認できたと考える．

高次元球の分析では，はじめに行ったような中間層の可視化ができないため，ヤコビアンの特異ベクトルを用いた分析を行うしかない．従って以降の高次元多様体を用いた検証では，ヤコビアンの分析による結果について記述する．

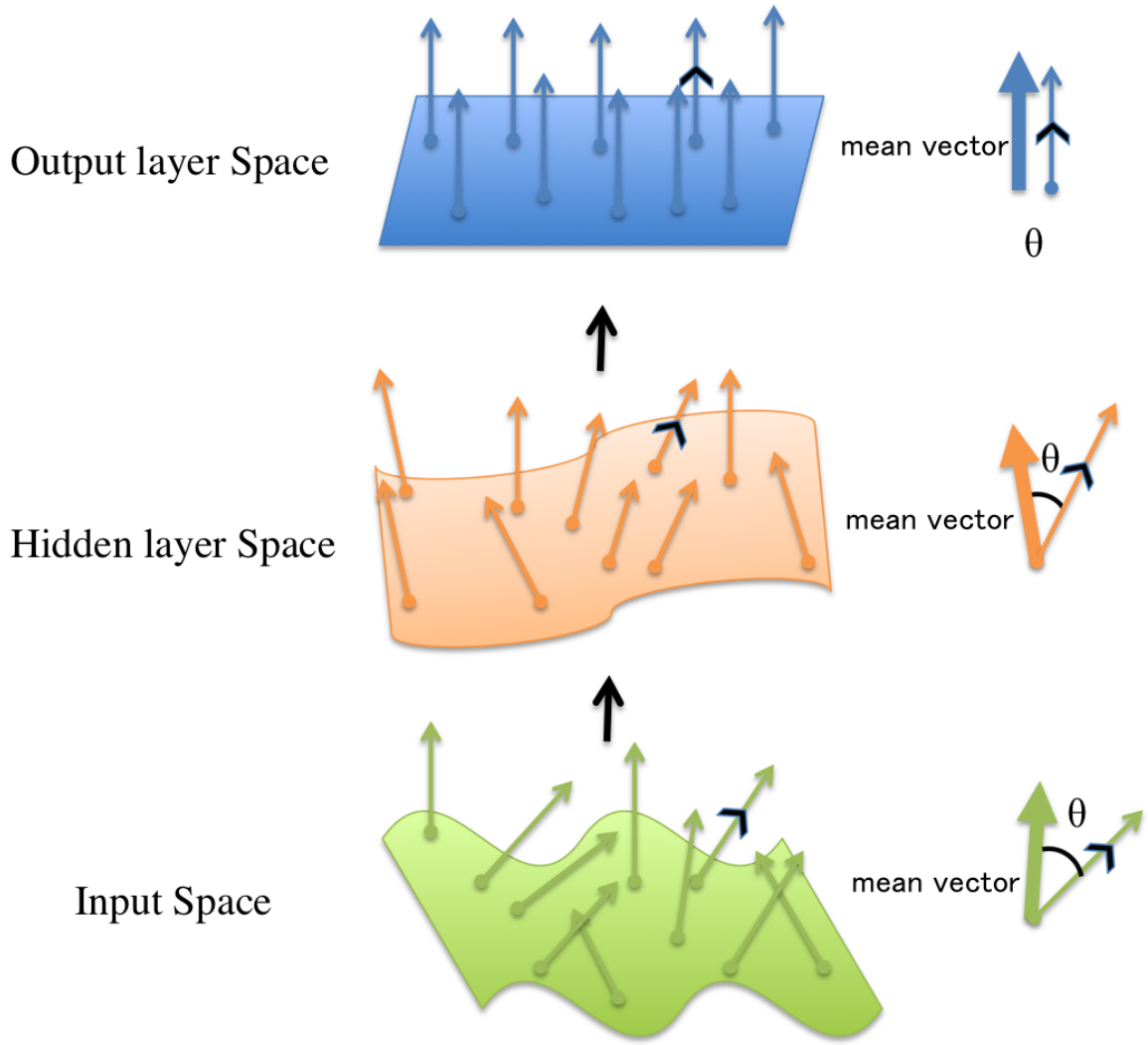

図 3.4: 左特異ベクトルによって算出される垂直ベクトルと平均ベクトルの偏角の定義
左特異ベクトルによって算出される垂直ベクトルとその平均ベクトルのなす角を偏角とする．図のように層をへるつどに多様体が展開され，最終的に大域的な空間が実現された場合，全ての垂直ベクトルの方向が平均ベクトルの方向と一致する（$\theta = 0$）．

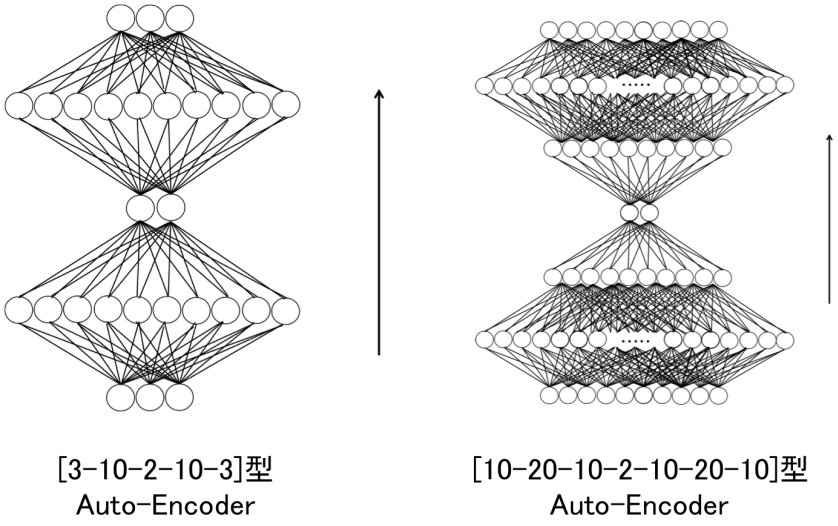

図 3.5: 使用した Deep Auto Encoder

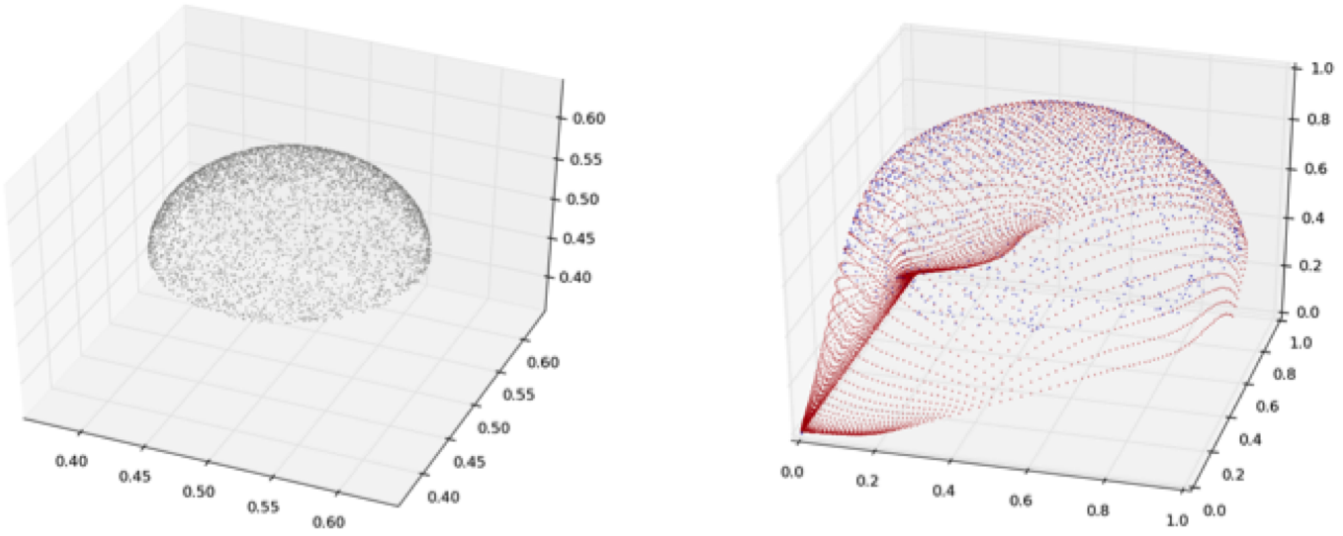

図 3.6: 生成モデルを用いた学習結果の可視化（3 次元入力）
左図：学習用データセットの分布，右図：学習結果（青点）と中間層（2 ノード）に形成された座標系のマッピング（赤点：中間層をグリッド状に刺激した結果）

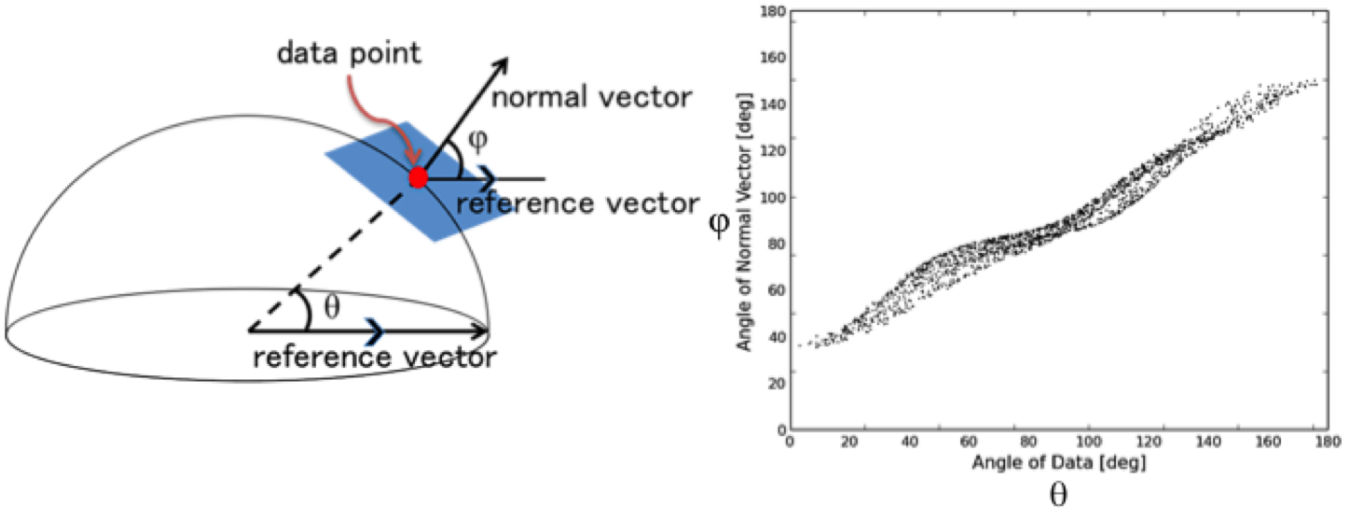

図 3.7: 右特異ベクトルと接空間の垂直ベクトル（3 次元入力）
左図：基準ベクトル（reference vector）と偏角（θ，φ）の定義，θ がデータセットの分布から決定される垂直方向．φ が，写像関数の分析から決定される垂直方向を表す．　右図：データから算出された垂直ベクトルの方位（θ）と右特異ベクトルから算出された垂直ベクトルの方位（φ）の比較．この二つが一致していれば ($\theta = \varphi$)，右特異ベクトルが多様体の接空間を捉えているといえるが，この実験では一致したなかった．これは，正方形で半球を捉える際に，正方形を伸縮させなければならないことに起因する．

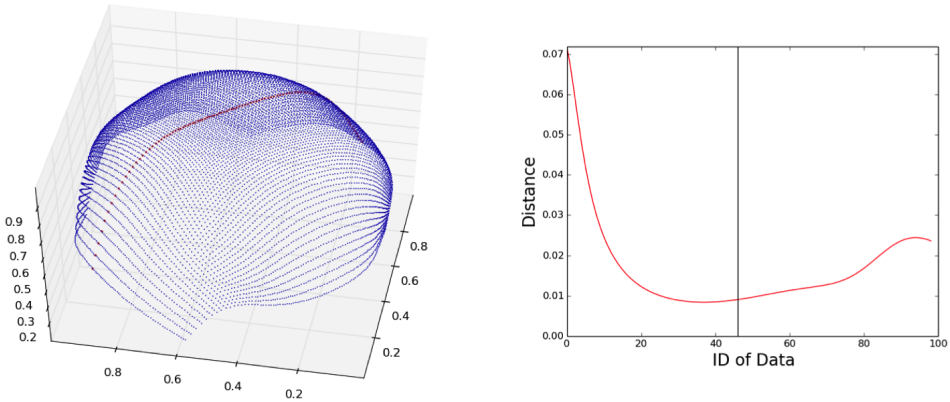

図 3.8: 中間層空間の多様体へのマッピングと座標系の伸縮
左図：図中の赤点にそって近傍点間の距離を算出，右図：近傍点間の距離をプロットしたもの．大域的な座標系が多様体上にマッピングされていれば，全ての点は均等に配置されるはずだが，半球の頂点付近（図中黒線）では密度が高く，半球の切断面付近では密度が低くなっていることがわかる．

3.3.2 実験結果

まず，ヤコビアンの特異値分布をみることで，DNN が多様体構造を捉えているかを検証した．その結果，ノード数が多様体の次元数 m である layer3 に向かうにつれて，多様体の次元と同じ数の特異値だけが値を持つような分布へおおよそ収束していくことが確認された（図 3.9）．また，大域的な座標系から入力空間の多様体へ逆に写像していく変換（layer3∼layer6：Decoder）の特異値分布をみると，その変換において，正確に多様体の次元が保存されていることが確認される（図 3.10）．この違いは，layer1∼layer3（Encoder）では，入力空間の圧縮による多様体の抽出と多様体の展開が同時に行われている一方，Encoder では多様体の展開のみを行えばよいという違いに起因すると考えられる．特に Encoder において多様体の次元と明確に一致する分布が得られていることから，DNN が多様体の次元をとらえていることが確認された．また，Decoder の高次層をみることでも，おおよその多様体の次元を知ることができることも確認できた．

次に，DNN が入力空間から多様体を抽出できているかをみるため，ヤコビアンの右特異ベクトルについて検討した．ここで，先ほどの 3 次元空間中の 3 次元球の場合と同様に偏角を定義する．ただし，今回は空間が 10 次元であるため，特異値が 0 となる特異ベクトルが複数存在する．そこで高次元球の分析では，特異値が 0 とならない特異ベクトル（多様体の接線方向ベクトル）との直交条件から多様体の垂直方向を推定した．この推定された垂直方向ベクトルと基準ベクトルとの偏角と，データセットから予想される垂直方向ベクトルと基準ベクトルとの偏角を比較すると，球の頂点付近（図の 90 °付近）においては，3 次元空間中の 3 次元球の場合と同様に $\varphi = \theta$ となることが確認された（図 3.11）．ただし，直交条件からだけでは，ベクトルの方向が正負逆になる自由度が残るため，分布には，$\varphi = -\theta$ となる直線も確認される．一方で，半球の切断面付近では 3 次元入力の場合と同様に $\varphi = \theta$ からのずれが観測された．このことから，この現象は多様体の次元によらず生じるものと考えられる．

以上から，n 次元正方形で一様に被覆できない多様体を観察する場合に誤差が生じる可能性があるものの，DNN はデータセットに多様体構造がある場合，その多様体の次元やおおよその形を知るための観測機として利用可能であると結論する．

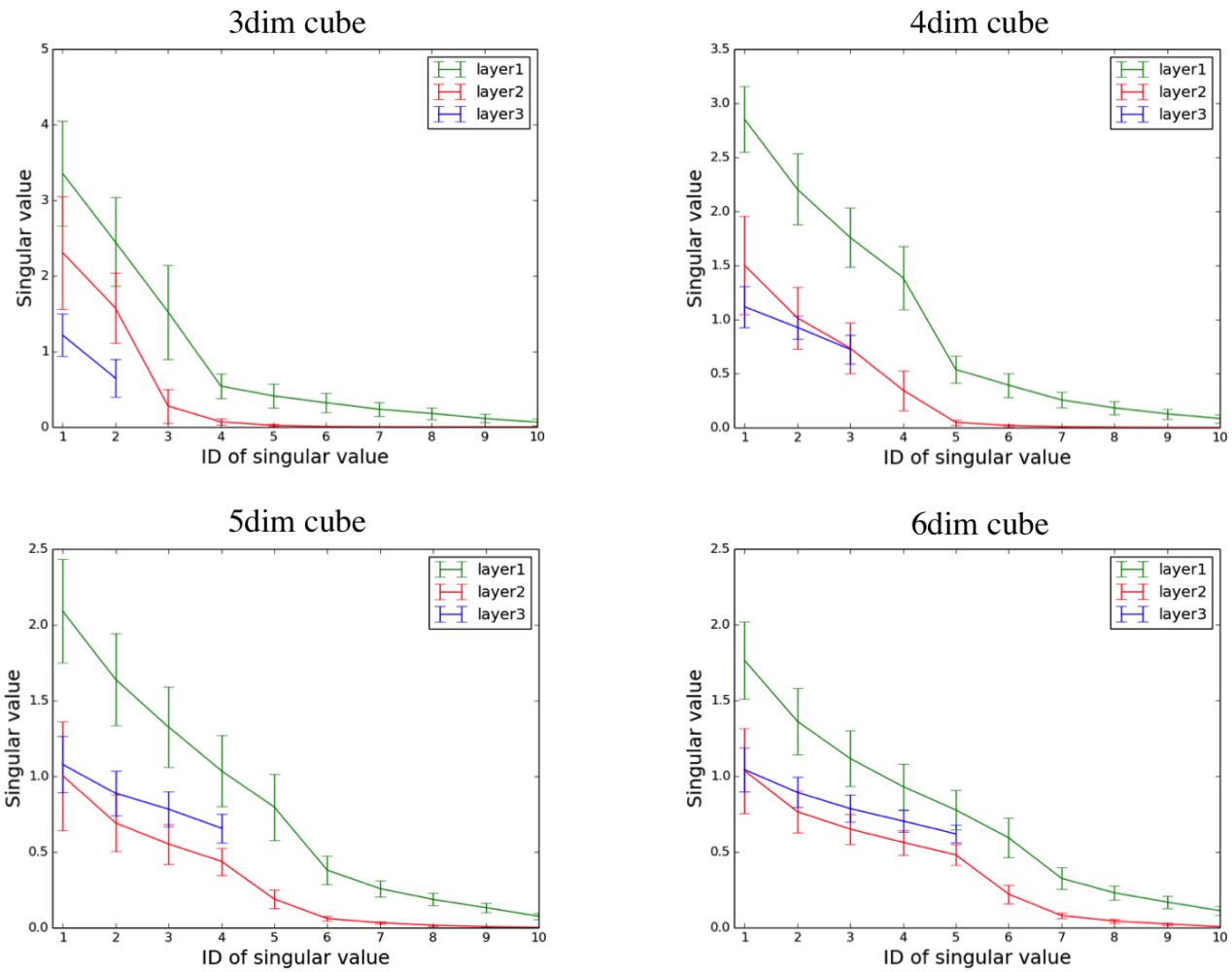

図 3.9: Encoder の特異値分布（10 次元入力）
3~6 次元球のどの結果でも，層をへるごとに多様体の次元と一致する特異値分布に収束している．

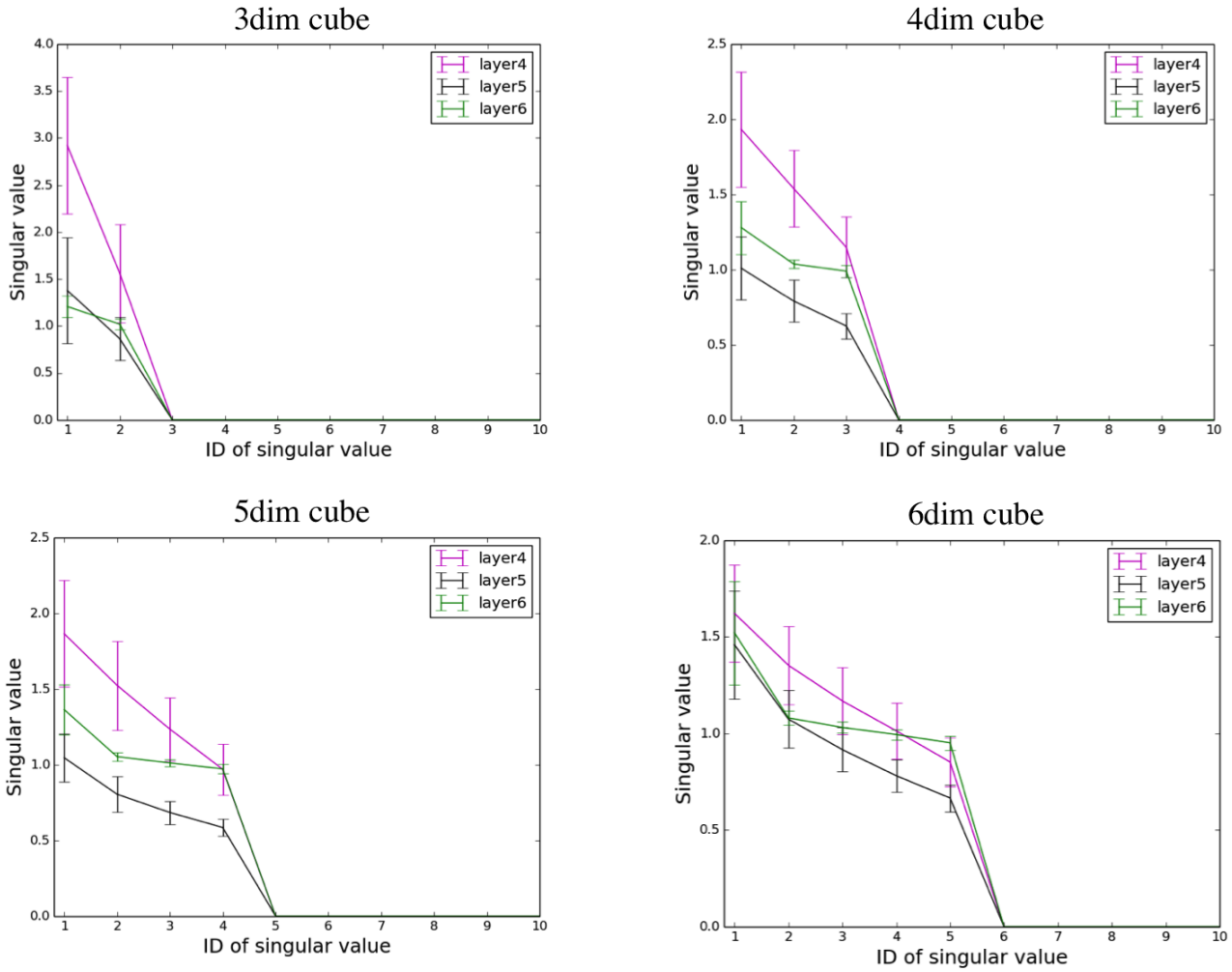

図 3.10: Decoder の特異値分布（10 次元入力）：生成モデル（down path）
Decoder 経路の特異値分布．次元を維持したまま入力を decode している．

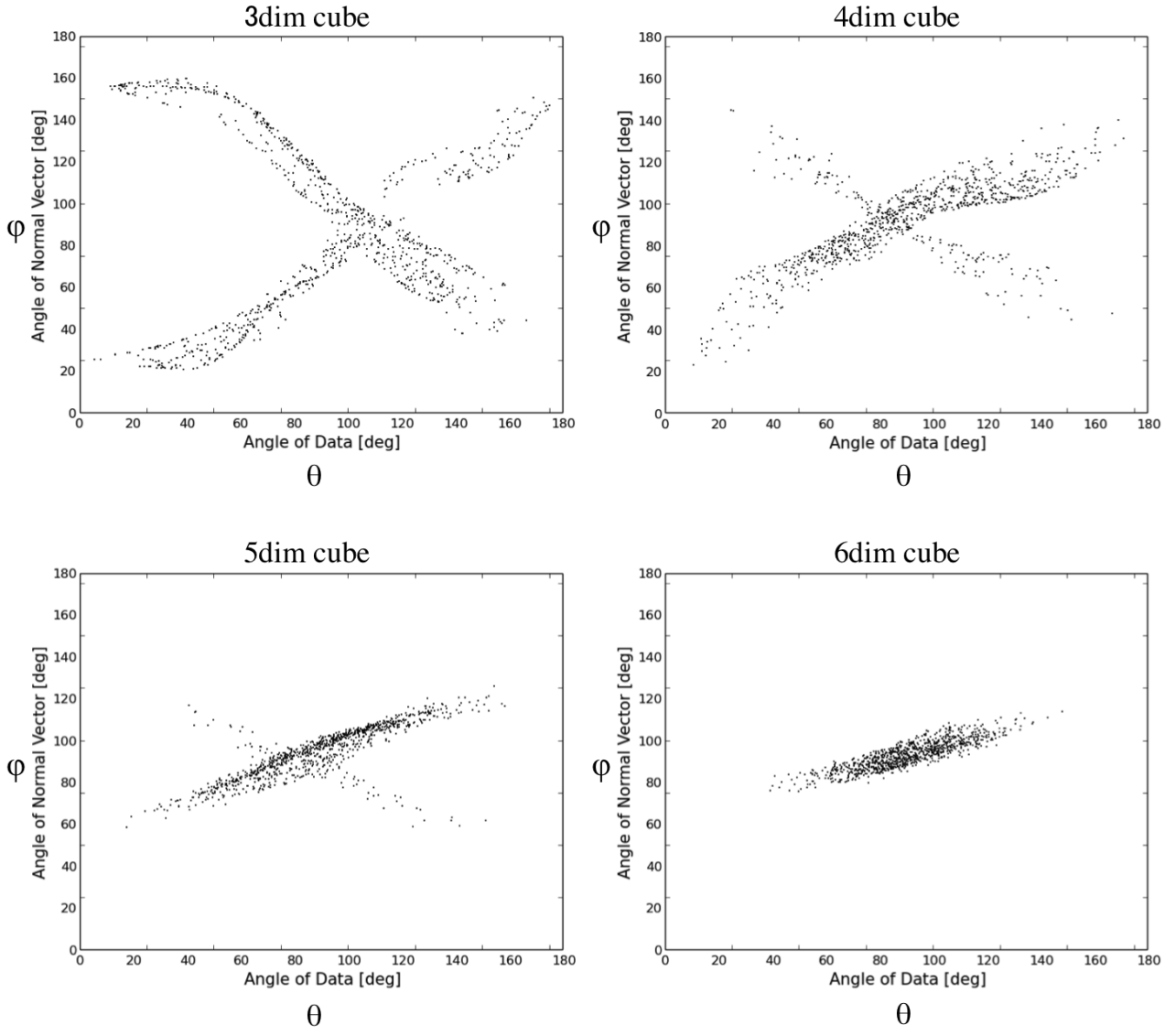

図 3.11: Input∼ Layer3 写像関数の右特異ベクトルと接空間の垂直ベクトル（10 次元入力）
先に定義した偏角（図 3.7）の分布．θ=φ となることから，3 次以上の多様体でも DNN が接空間を捉えられることを示している．

3.4 DNN観測装置の性能検証2：MNIST

前の人工データを用いた実験によって，DNNがデータセットの幾何構造を観測する手段として有用であることが確認された．本節ではさらに人工的でない実践的なデータセットに対してもDNNが観測手段として有効であるかを検証する．その為に，MNIST[37] 手書き文字データセットでトレーニングされたDNNの分析を行った．前章でみたように，多様体学習 [59] や特殊なDNN[30] [15] を用いた研究によって，MNISTは，多様体構造を持つこととそれが$O(1)$程度の次元であることが示唆されているデータセットである．特殊なDNNとは前の実験で使用したような中間層の次元を2次元に絞ったDeep Auto Encoderのことである．一方で，実用上このような極端に中間層のノード数を制限したネットワークを使用することは稀である．ほとんどのデータ分析では多様体の次元が不明な為である．そこで本実験ではより実践的に中間層の数を大きくとったDNNを用いる．その上で，これまでに判明しているMNISTデータセットの構造を観測できるかを検証した．

3.4.1 実験方法

本実験では，Deep Belief Networks[29]（以下，DBN）モデルを用いる．DBNは多数の層からなるニュー ラルネットワークであり，各層はRestricted Boltzmann Machine（以下，RBM）によって構成される．RBMは，層状にネットワーク構造が制限されたボルツマンマシンのことである（付録A.2参照）．学習はAuto Encoderと同様に入力と出力を一致させるように行う．ただし学習の指標としては，Auto Encoderで用いたような二乗誤差関数ではなく，入力データの確率分布とネットワークの出力の確率分布の分布間距離をあらわすKL-divergenceを用いる．

実際に，Gaussian-binary型RBMのエネルギー関数は，前に用いたAuto Encoderの一種であるDenoising Auto Encoder[63] のエネルギー関数（score matching[33] で確率分布を推定し定義）と等価になることが示されている [62].（ただし，エネルギー関数全体にかかる係数項に差異がある.）Denoising Auto Encoderとは，入力次元の一部を取り除いた状況で学習を行うAuto Encoderのことで，その結果として，欠損したデータから元のデータを復元するような写像関数が獲得される．入力次元の一部を取り除くことはちょうど，多様体の周辺にデータを再配置することに対応すると考えられ，実際にContractive-Auto-Encoderと類似した特異値分布が獲得されることが実験的に示されている [48].

このように，DBNはDeep Auto Encoderと類似しており，同様にデータセットの観測装置として利用可能であることが強く期待される．

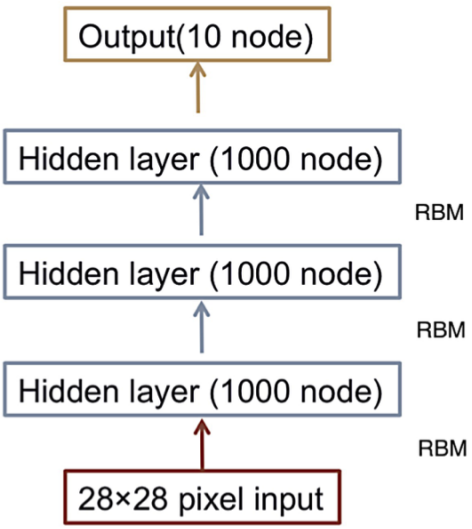

図 3.12: 使用した Deep Belief Networks

また，DBN は，教師なし学習を用いた最初期の深層学習アルゴリズムであり，教師なし学習によって pre-training された DNN を用いることで，それまで最も高いパフォーマンスを示していたアルゴリズムの 1 つであった SVM（Support Vector Machine）と同程度の認識精度を，一部のラベルデータだけを用いた fine-tuning（教師あり学習）だけで実現している．

以上のことを踏まえ，本研究では MNIST データセットの観察装置として DBN を用いた．具体的には，図 3.12 にあるような [28*28 - 1000 - 1000 - 1000 - 10] となる DBN を構築し使用した．全体を最適化する fine-tuning は，ラベルデータを元にした教師あり学習（Back Propagation）によって行った．

観測するデータセットである MNIST[37] は，人が手書きで書いた 0~9 までの数字の画像データ 70,000 サンプルによって構成されており，それぞれの画像に対して 0~9 のうちのいずれかのラベルが付与されている．fine-tuning 後の MNIST データセットの test error は，1.26%であった．また，学習率や重みの初期値などの使用した全てのハイパーパラメータは，theano ライブラリの DBN チュートリアル [55] のパラメータと一致させてある．

3.4.2 実験結果

pre-training（教師なし学習）後と fine-tuning（教師あり学習）後について，それぞれ特異値分布と特異ベクトルを計算した．

まず，pre-training 後の特異値分布をみると，層をのぼるにつれて勾配が急峻になる特異値分布がみられた（図 3.13 の左上）．人工データでの分析でみたように，特異値分布は，多様体の次元数の前後で急激に 0 に減少する傾向があった．本実験の結果はこれに類似している．また，0 でない特異値の数は $O(1)$ 程度であった．これは先行研究の結果と整合性がある結果である [59][30] [15].

次に，右特異ベクトル（図 3.13 の左下）をみると，大きな特異値に対応する特異ベクトルにおいて，入力データに対応するよな構造をもったベクトルが見られる．これは，MNIST のデータ分布が部分ごとに違う接空間をもつことを，つまり多様体構造を持っていることを示唆する．またそのベクトルの様相は，第 2 章で論じたような多様体が回転や平行移動に対応する普遍性によって形成される接線ベクトルと類似している（図 3.14）．図 3.14 は，手書き数字を 5 °回転した場合のデータと元のデータの差分（多様体の接線方向）を可視化したものである．この結果は，今回捉えた多様体もデータ点まわりのなんらかの変換に対する普遍性と関係することを示唆する．一方，1 より非常に小さい特異値に対応する右特異ベクトルをみると，ノイズのような大域的な構造をしており，前に説明した力学的な視点で考えると，ノイズ情報を圧縮して何らかの普遍性に対応する feature 情報を保存するという DNN 内の情報の流れが想起される．これは，入力空間を圧縮しながら多様体を見つけ出すという，仮定した DNN の機能の 1 つが存在することを支持する．

次に，fine-tuning 後の特異値分布，右特異ベクトルをみると（図 3.13 の右側），特異値分布全体が pre-training 後と比較して増加し，特異ベクトルの形も変わっていることがわかる．

このような現象が生じるのは，fine-tuning（教師あり学習）が教師なし学習とは違うダイナミクスをもつことが原因であると考えられる．

つまり，データをラベルごとに正しく分類することをコスト関数としてトップダウン方向に Back-Propagation を行う fine-tuning の際には，pre-training 時のような教師なし学習による多様体構造の抽出と展開ではなく，クラス間に分離面を引くようなことを学習していると考えられる．特に特異値分布全体が増加することは，ネットワークを流れる情報全体が増加していることを意味しており，このことは，fine-tuning 時に文字画像データセットそのものにないような，どの数字なのか微妙な形をした文字が，どのクラスに所属するかといったラベル情報が与えられた結果と考えることができる．このような機構の存在によって写像関数が変化した結果，特異ベクトルにも pre-training 時との間に差分が生まれたものと考えられる．

本実験のネットワークは，人工データの学習時と違い，最もノード数の少ない中間層でも 1,000 次元もある．そのため，人工データでの実験のように中間層にお

いて左特異ベクトルから算出される多様体の垂直方向が，全てそろう保証（大域的な座標系になっている保証）がない．そこで，次にこれを検証する．

ヤコビアンの左特異ベクトルから，次のような手順で多様体の垂直方向成分のベクトルを算出した．前に示した特異値分布から各層で捉えている多様体構造の次元 n を推定した．具体的には，layer1 を 50 次元，layer2 を 10 次元，layer3 を 2 次元とした．その上で，特異値が大きい順番に次元数分の左特異ベクトルを選び，これとの直交条件から垂直方向ベクトルを算出した．

一方，1,000 次元空間中で直交条件式が 2 つしかないため，垂直方向の自由度が (1,000 - 多様体の次元数) 個残ることになる．

この自由度の中から垂直ベクトルとして，$(1, \cdots, 1, x_i, x_{i+1}, \cdots, x_{i+n}, 1, \cdots, 1)$ となるベクトルを 1000-n 個（$i = 0 \sim 1000 - n$）算出し，その平均値の振る舞いを検討した．ここで，$x_i, x_{i+1}, \cdots, x_{i+n}$ は直行条件から算出されたベクトルの要素を意味する．算出した垂直ベクトルは互いに独立でないため，この分析で確認できることは正確には大域的な座標系の獲得の必要条件のみになるが，方向の違う 1000-n 個の垂直ベクトルをみる為，十分に確認できると考えた．この算出法のもと，左特異ベクトルの垂直方向成分ベクトルを算出し，実験 1 で説明した方法で偏角を計算したところ（図 3.4），layer3 において全ての垂直方向ベクトルが同じ方向を持つようになることが確認された（図 3.15）．このことから，DNN が大域的な座標系への写像を行っていることが確認されたと考える．

ところで，0~9 のそれぞれのクラスがそれぞれ 1 つの部分多様体に対応する場合，微分多様体の定義上，そのクラスの中での次元は一致しなければならない．従って，クラス間で多様体の次元が違う場合，クラス毎に特異値分布が違う可能性がある．また，多様体の曲がり具合や多様体上での接空間の伸縮などの，幾何学的な構造にもクラス間で相違があれば，その影響を特異値分布が受けるとも考えられる．そこでここで，クラス毎に平均化した特異値分布をみてみることにする．すると，第 2 層目や第 3 層目において，クラス毎に大きく異なる特異値分布がみられた（ 図 3.16）．これをより客観的にみるために，784 個ある特異値の値を 1 つのベクトルとして主成分分析を行った．この結果，特に第 2 層目以降で，クラス毎にクラスターが形成されることが確認された（ 図 3.17）．特異値分布は，多様体の局所的幾何構造の情報のみしかもたないにも関わらず，このようなクラスタが形成されることは，データセットの幾何学的構造とデータセットのクラスとの間に強い関係性があることを示唆すると考えられる．

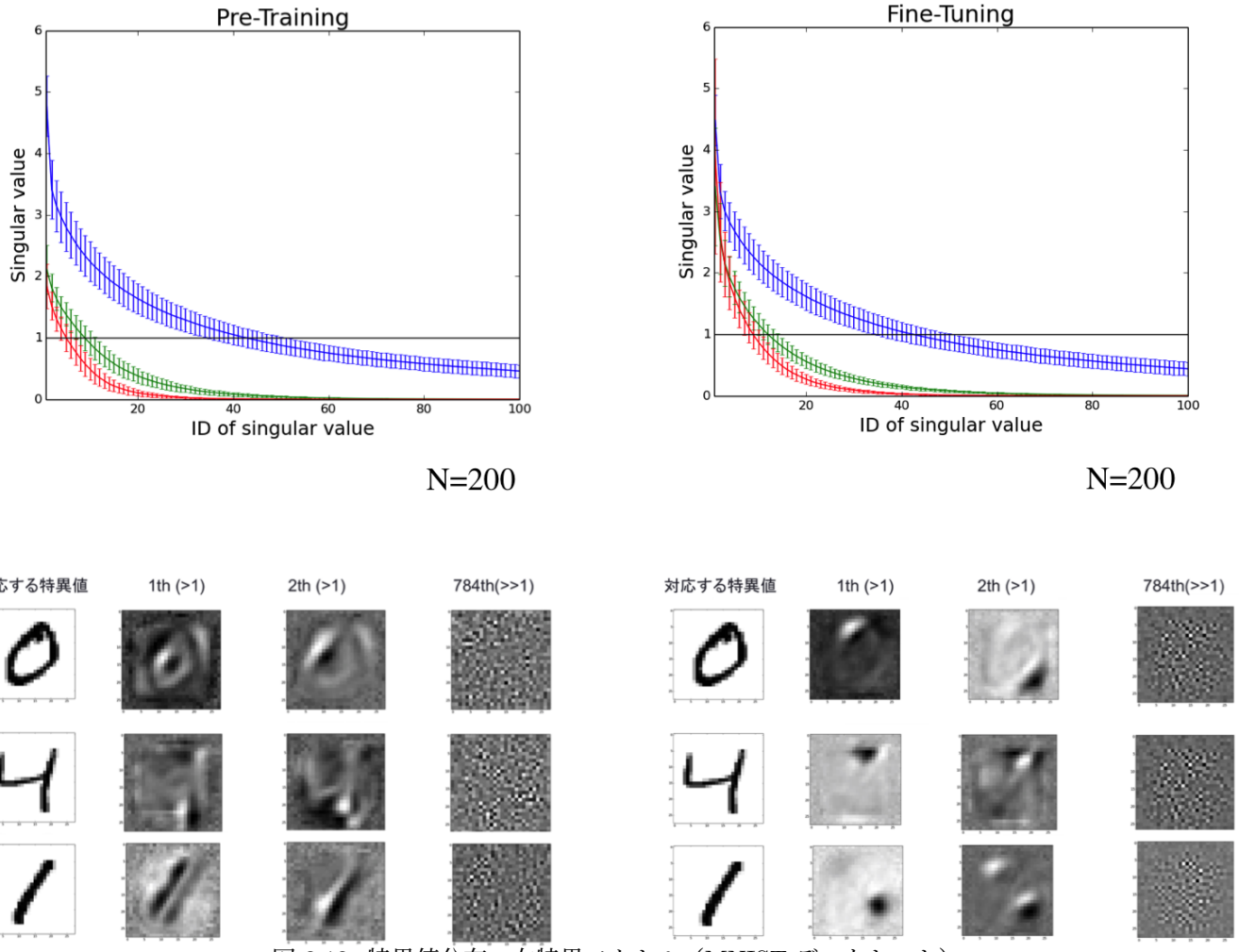

図 3.13: 特異値分布・右特異ベクトル（MNIST データセット）
上段左：pre-training 後の特異値分布．層をのぼるごとに，特異値分布が急峻になっていることがわかる．，上段右：fine-tuning 後の特異値分布．pre-training 後に比べ，特異値が 1 以上である特異値の数が増加している．教師データによって，ネットワークで保存される情報が増大していると予想される．下段左：pre-training 後の右特異ベクトル．図 3.14 のようなデータに対応したベクトルがみられる．下段右：fine-tuning 後の右特異ベクトル．pre-training 後と違う様相になっており，多様体とは違うものも捉えている可能性がある．エラーバーは標準偏差をあらわす．

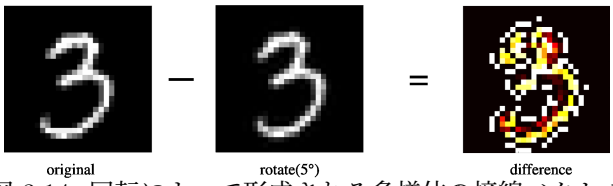

図 3.14: 回転によって形成される多様体の接線ベクトル
画像の回転によって多様体が形成されているとした場合の接線ベクトル．5 °回転した画像との差分から，接線ベクトルを算出.

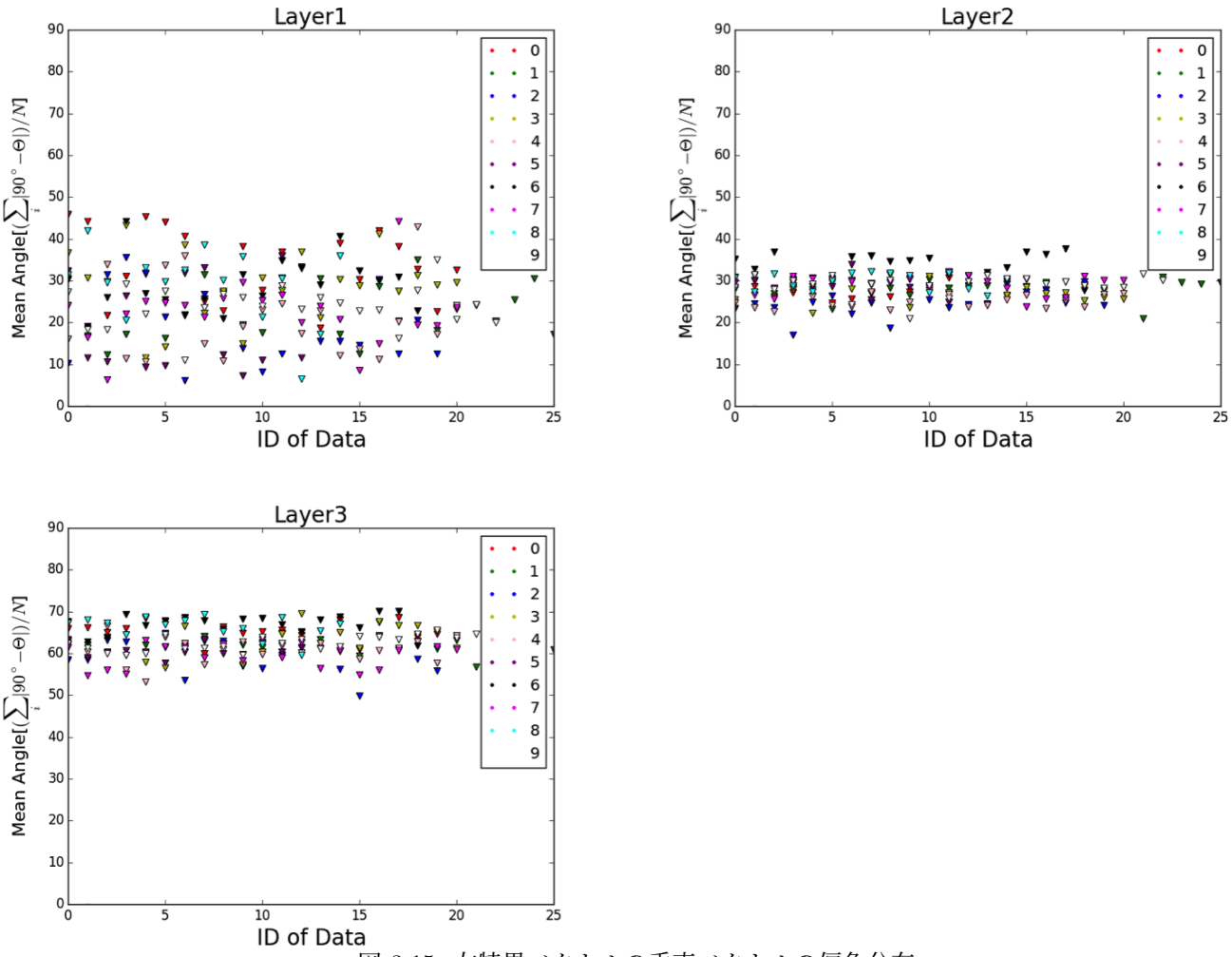

図 3.15: 左特異ベクトルの垂直ベクトルの偏角分布

第 1，第 2 左特異ベクトルとの直交条件によって算出されたある垂直ベクトルの偏角を θ とする．ベクトルの存在する次元に比べて直交条件が少ないため，独立な垂直ベクトルが複数定義される．クラス毎に垂直ベクトルの平均を求めた場合，それと各データ点で定義される全ての垂直ベクトルの偏角が $\theta = 0 or 180$ °となった場合，多様体の接空間が大域的な座標系になっていることになる．そこで，それらの偏角と 90 °との差分の絶対値の平均値を算出しプロットした．全ての垂直ベクトルが同じ方向を向く場合，この値は 90 °になるはずである．これをみたところ，第 3 層において，相対的に偏角が $\theta = 90$ 付近に近づいていくことが確認された．これは，大域的な座標系が実現されている傾向があることを示唆する．

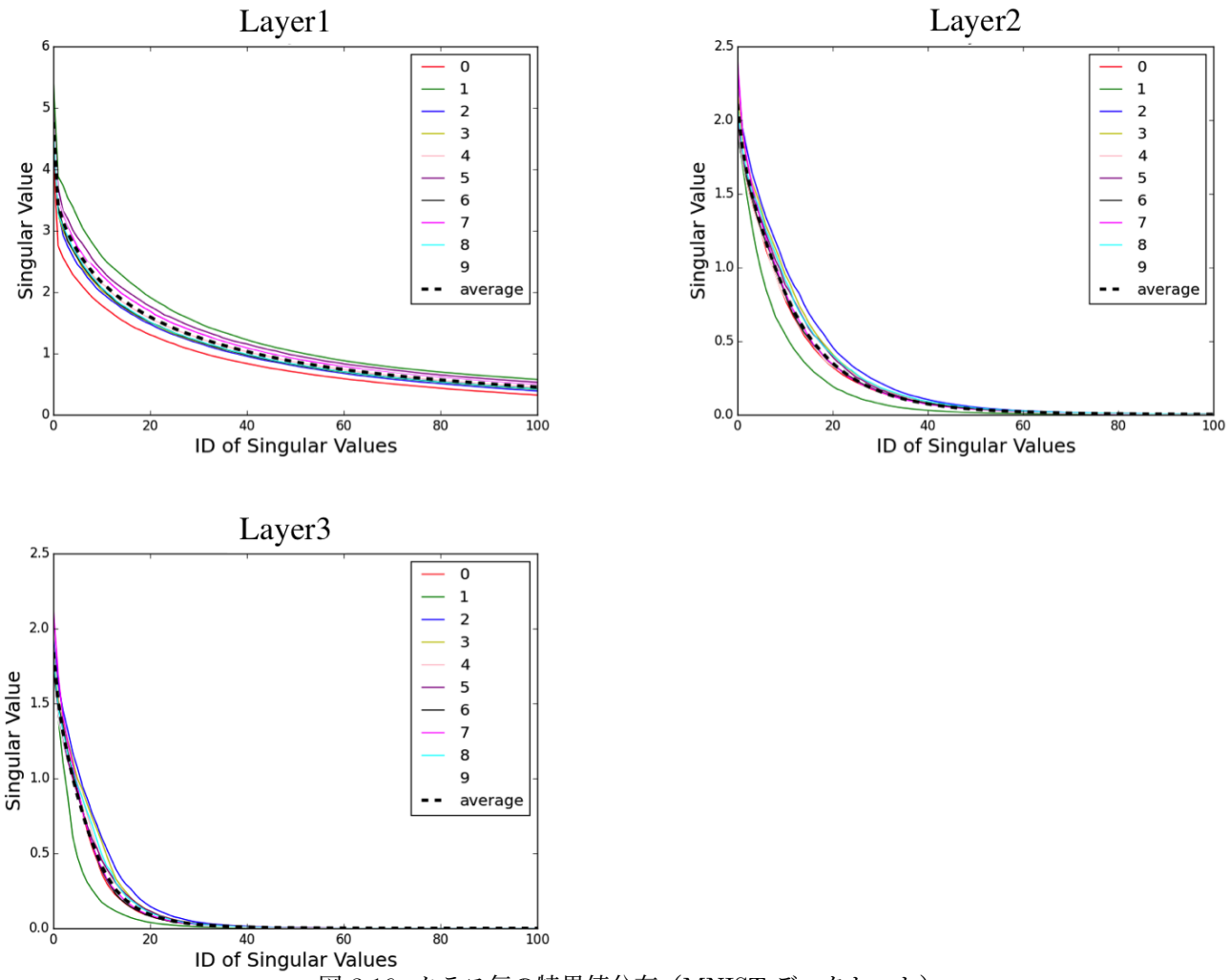

図 3.16: クラス毎の特異値分布（MNIST データセット）
0~9 の手書き文字毎の特異値分布．特に ”1 ”が顕著であるが，層をへる都度，クラス毎に違う分布へ収束していくとわかる．

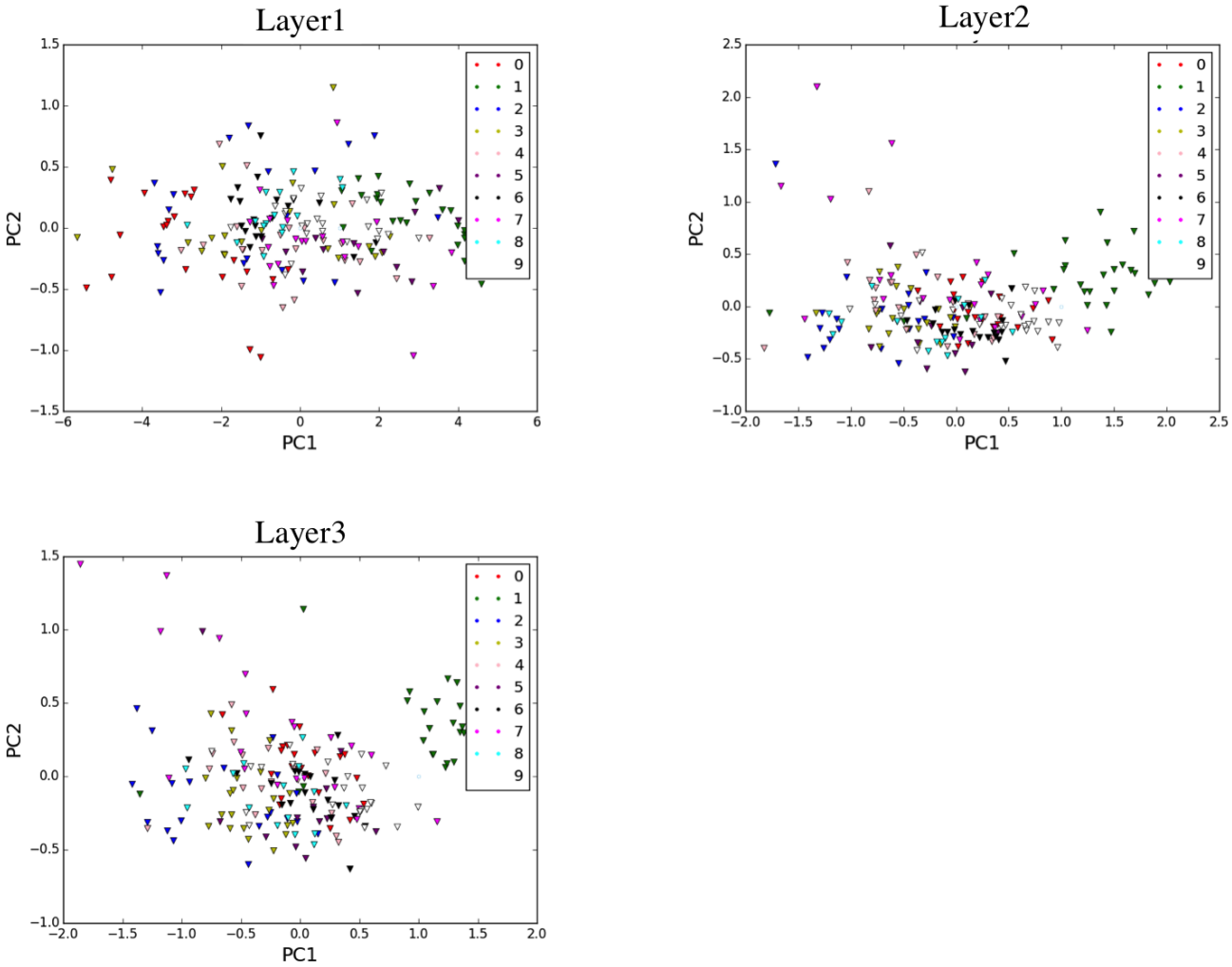

図 3.17: 特異値の主成分分析（MNIST データセット）
784 個ある特異値をインプットベクトルとし，主成分分析で 2 次元へ圧縮した結果．高次の層で，クラス毎にクラスタを形成していることがわかる．

3.5 まとめと考察

本章では，DNN をデータの幾何構造を観察する為の観測装置として使用することを目的に，その性能・性質を調べることを行った．

具体的には，多様体仮説

- データセットは，そのデータの次元 d より十分に低い次元 d_M を持つ（微分可能）多様体上に埋め込まれれている．

を満たすと考えられるデータセットを用いて，

- パフォーマンスの高い学習済み Deep Neural Networks（以下，DNN）は，上の多様体を多様体と同じ次元の大域的な座標系へ写像する機能をもつ．

という仮説の検証を行った．

具体的には，多様体と同じ次元の中間層を持つ Deep Auto Encoder に 10 次元空間中の n 次元球を学習させその写像関数を分析した．その結果，DNN の写像関数のヤコビアンの特異値分布の分析によって DNN が多様体の次元を捉えていることを，右特異ベクトルの分析によって DNN が接空間をおおよそ捉えていることを確認した．一方で，本研究でデータセットの構造として設定した半球のような，多様体と同じ次元の正方形で引き伸ばしなしに被覆できない構造がある場合，右特異ベクトルの方向に誤差が生じることも確認された．データセットの観測に DNN を用いる場合は，この誤差の存在を常に意識する必要がある．

これまでの研究では，可視化に頼った方法によってこの仮説を確認しており [30][71]，定量的な手法による検証は本研究が初めてであった．またこれと関連して，可視化によって検証が行えない高次元空間中の n($n \geq 3$) 次元多様体での検証を行ったことも本研究の貢献となる．また，前述したような半球の切断面付近での誤差は入江らの研究 [71] でも確認されていたものの，入江らはこれを無視していた．本研究ではその誤差について検討し，前述したようにこの誤差の原因として幾何学的な要因を提示することができた．この現象が生じる根底には，DNN の写像関数が同相的な変換をしていることがあると考えられる．従って本研究で判明した知見を応用し，正方形では伸縮なしに被覆できないような多様体構造を持つデータセットを学習する場合に，それを正方形で自然に被覆できるよう，あえて DNN 写像関数の同相性を壊すような機構を導入することで，よりパフォーマンスの高いアルゴリズムが開発できる可能性が示唆されたと考える．今後，この点も検討していきたい．

次に本研究では，実践的なデータセットである MNIST データセットによってトレーニングされた DNN も仮説を満たす写像機能を持つかについて，定量的な検証

を行った．MNIST データセットは，hinton らの研究 [30] 等によって多様体構造の存在と，その次元が $O(1)$ 程度になることが示唆されている．一方で，これらの研究は中間層が 2 となる Deep Auto Encoder を用いることで，可視化を通して仮説を検証しており，十分な検証がなされているとは言い難い．また，MNIST の識別においてパフォーマンスが高い DNN では，このような極端に小さな中間層をもつネットワーク構造は用いられていない．そこで本研究では実際に使用されるような中間層がより高次元な DNN を定量的に分析し，実践的なデータセットにおいてもパフォーマンスの高い DNN が仮説を満たすことを確認することを試みた．

その結果，pre-training（教師なし学習）後の特異値分布の分析から DNN が捉えている多様体の次元が $O(1)$ になることが確認され，右特異ベクトルの分析から，DNN が多様体の接空間を捉えていることが，左特異ベクトルの分析から DNN がそれを大域的な座標系へ展開していることが確認された．

さらに，fine-tuning（教師あり学習）後の特異値分布の分析によって，fine-tuning（教師あり学習）が教師なし学習とは違うダイナミクスをもつことがわかった．特に fine-tuning 後の特異値分布が全体的に増大することから，fine-tuning によって DNN に情報が追加されていると考えられる．これらの結果は，pre-training 時のような教師なし学習の際には多様体構造を大域的な座標系へ変換するような単射的（n 対 n 写像）な写像機能が獲得され，fine-tuning 時のような教師あり学習では，違うクラスの部分多様体の間に存在するような判別の難しいデータがどのクラスに所属するかという情報を与えるような，クラス間に分離面を引くような写像機能（n 対 m 写像：m<<n）が獲得されると描像が考えられる．一方で，fine-tuning 後の特異値分布や右特異ベクトルが，pre-training 後と比較して大きく違うものとはなっていないこと，及び，pre-training 法の導入によって初めて初期の深層学習が可能となったという事実から，教師あり学習時においても多様体構造の情報は有用に機能するものと考えられる．つまり，DNN は，これら 2 種類のダイナミクスの相互作用によって写像を行っていると考えられる．このことは，次章でみるような pre-training を用いない DNN においても，多様体構造を大域的な座標系へ写像するような機能が獲得される可能性を示唆する．

また，クラス毎の特異値分布の形状の違いを PCA を用いた次元圧縮を通して検討したところ，特異値分布がクラス毎にそれぞれ違う形状をもつことが確認された．このことは，データセットの意味的なクラスと幾何的な構造が関係することを示唆しており，MNIST データセットが多様体仮説を満たすことを強く支持する．幾何構造と意味的なクラスの間の関係性をこのように直接的に示した研究はこれまでにない．

第4章　Deep Neural Networksによるデータセットの観測

4.1　DNNによるImageNetデータセットの観察

4.1.1　目的

この章の目的は，階層的な意味的カテゴリ構造を持った大規模な画像データセットに，どのような幾何的な構造があるかを観測することである．そのために本実験では，ImageNet データセット [35] を学習データとして用いた（図 4.1 にデータの例）．ImageNet データセットは，単語の定義や同義語のグループとグループ間の関係性が記述された英語の概念辞書（意味辞書）である WordNet[24] のオントロジーに従って，各単語（名詞）に対応する画像を収集したもので，現在約 1,500 万枚の画像が登録されている．

4.1.2　実験方法

ImageNet は，非常に大きな画像サイズのデータセットであり，それを学習するネットワークも大規模になる．本実験では，2012 年の画像認識コンテストに優勝した， Krizhevsky らによって開発された，畳み込み（付録 C 参照）や pooling（付録 C 参照），drop out（付録 D 参照）[53] 等の技術を組み込んだ DNN（AlexNet[35]）を分析対象とした．AlexNet は，教師あり学習（Back Propagation）によって学習されるが，Yosinski らは AlexNet の畳み込み層がデータの詳細によらず学習されることを示唆している [66]．また，Convoution Neural Networks は，画像データの並進に対する普遍性を取得するネットワークだと考えられており（付録 C 参照），これは多様体仮説に対応する．そこで本研究では，これらの層は多様体構造を抽出し，さらにそれを大域的な座標系へ展開する機能を持つと仮定した．従って，データの観察と同時にこの仮定の妥当性についての検証も行った．

具体的なネットワークの構築と分析には，公開されている学習済みパラメータデータ（DeCAF [22]）を用い，特異値・特異ベクトルは，120 枚の違う入力画像

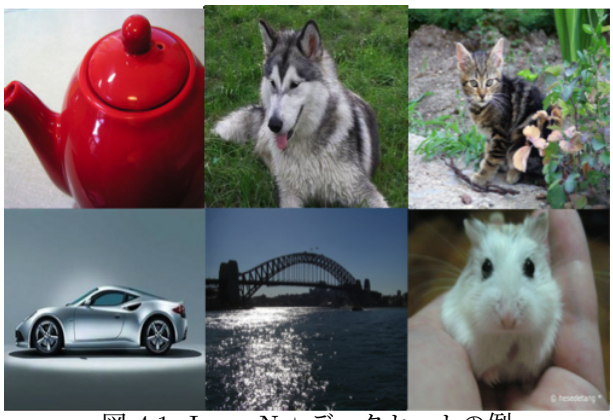

図 4.1: ImageNet データセットの例

に対して大きい方から順に 500 番目まで計算した.（行列のサイズが大きくなるため，Halko らによって開発されたアルゴリズムに基づいて，スパース行列でヤコビアンが表現された状態で SVD を行なった [28].）

また，このネットワークにおける活性化関数 $f(x)$ は，下式で定義される ReLU（rectified linear unit）関数を用いた.

$$f(x) = \begin{cases} x & (x \geq 0) \quad (4.1a) \\ 0 & (x < 0) \quad (4.1b) \end{cases}$$

AlexNet は，マルチ GPU での実行を実現するために，Convolution 層が 2 つのパスに分かれており（図 4.2 参照），この構造の結果として，それぞれのパスが自然に「色」と「形」に対応することが報告されている [35]．本実験で用いたネットワークにおいても，第 1 層の重みベクトルをみると，図 4.3 のようにそれぞれのパスが「色」と「形」に対応していることがわかる.

計算量の低減のため，本研究ではこの 2 つのパスのうち「形」を担当するパスについてのみの分析を行った.

4.1.3 結果と考察

計算の結果，高次の層において，少数の大きな特異値と，大多数のほぼ 0 の特異値という，急峻な特異値分布が見られた（図 4.4 参照）．また，特異ベクトルをみた結果，特異値の大きいベクトルでは，各データに固有のようにみえる，データの feature を捉えるベクトルがみられる一方で，特異値の小さなベクトルでは，空間的に広く分布したノイズのような構造がみられた．これは MNIST でみられたも

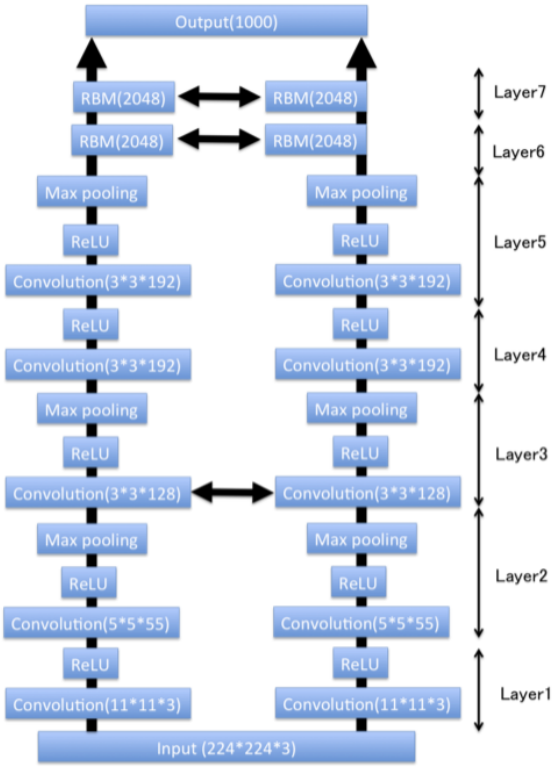

図 4.2: AlexNet のネットワーク

CNN（convolution neural networks networks）と pooling 層，ReLU 層を多層に重ねた上で RBM を追加した形状となっている．本研究では，図のように CNN 層，pooling 層，ReLU 層をまとめて 1 つの層と定義する．

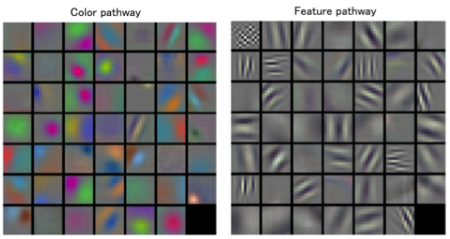

図 4.3: AlexNet の第 1 層目ウェイトマトリクスの可視化

ウェイトマトリクスは，対応する隠れ層ニューロンが担当する入力空間の部位を表すので，この図から，左経路が色経路，右経路が形経路を表すことがわかる．(図左側を左経路，図右側を右経路とする)

のと同様の構造である（図 4.5 参照）．これらの結果は，この DNN が多様体の接空間の間の写像関数を獲得していることと，ImageNet データセットが多様体構造をもつことを示唆する．これ以外の右特異ベクトルの例を付録 E に添付する．

このように，これらの結果は多様体の存在を示唆するものの，第 5 層においてもまだ 100 次元以上の次元が残っており，特異ベクトルの偏角による分析は難しい（次元が高いとわずかなズレで容易に内積が 0 となってしまう）．そこで，DNN が多様体の接空間の写像を行っていることを他の方法で確認することを試みた．具体的には，以下のようなことを行った．もし，1 より大きい特異値に対応する特異ベクトルが多様体の接空間を表しているとすれば，入力データにその方向の摂動を加えたとしてもニューラルネットワークの出力はあまり変動しないはずである（クラスが変わらないため）．一方，多様体の垂直方向である 1 より小さい特異値に対応する特異ベクトル方向への摂動を加えると出力は大きく変動するはずである．これを確かめたところ，実際にニューラルネットワークの出力が，接線方向の摂動に対して相対的にロバストであることが確認された（図 4.6）．このことは，ImageNet データセットが多様体構造をもっていること，及び AlexNet がその接空間を入力空間から抽出するような写像関数を学習していることを意味する．ちなみに，小さな摂動に対しては摂動の方向によらずほぼ同じ出力変動となっているが，これは前に説明したように，多様体仮説はデータセットの分布が多様体 ”周辺 ”に分布するとしており，DNN はこれらを多様体に埋め込むような写像を行っていると考えられる．この吸引領域がこの小さな摂動の領域に対応すると考えられる．また，接線方向への摂動にもかかわらず大きな摂動をかけた場合に出力が大きく変化しているが，これは多様体が曲がっていることに起因すると考えられる．この事実を用いると，この分析法から多様体の曲率情報を得られる可能性が考えられる．

次に，多様体から離れるような摂動と特異値分布の関係を調べた．そのために入力画像にノイズを付加した上で，特異値の算出を行なったところ，高次の層において，ノイズの増大に応じて特異値分布全体が小さくなっていくことが観察された（図 4.7 参照）．また，このときの出力は不正解となっている．これは，データ点が多様体から大きく離れた場合，DNN 内でうまく情報が伝達されなくなることを意味していると考えられる．この結果より，学習がうまくいかない状態を特異値分布で捉えることができる可能性があると考えられ，次章で説明する DNN の大自由度力学系への応用では，この特異値分布とパフォーマンスの関係を用いたハイパーパラメータチューニングに挑戦している．

最後に，前章の MNIST データセットを用いた分析と同様にクラス毎の特異値分布をみた．今回クラスとしてとらえたのは，図 4.8 の凡例にあるものである．それぞれ右が ImageNet データセットにおいて最も低次のカテゴリになっており，左

が高次のカテゴリとなっている．カテゴリ毎の特異値分布をみると，層を通過するにつれてカテゴリ毎に特異値分布が同じ分布の形に縮退していくようにみえた（図 4.8）．また，MNIST の場合と同様にして特異値をベクトルとみなして主成分分析を行うと，同様に層を経るにつれてクラス毎にクラスタを作る傾向があるようにみえた．特に，低次のクラス（同色・同形の点）は，より低次の層でクラスタを形成しているようにみえた（図 4.9）．

これらの主観的な観察を踏まえ，この結果を定量的に検証した（図 4.10，図 4.11）．具体的には，良いクラスタの指標として使用される，クラスタの大きさを表す標準偏差（より良いクラスタは狭い領域に分布する）と，クラスタ間の平均距離（より良いクラスタはお互いに離れている）を用いて上の観察結果を検証した．ただし，layer 間でのクラスタ形成傾向の比較を行うため，指標は各層のデータ分布全体の標準偏差で規格化した．具体的な指標の算出は次の通りに行った．まず，低次のクラスと高次のクラスについてそれぞれクラス毎に前述した 2 つの指標の算出を行った（図 4.10，図 4.11 の赤）．これと，同じデータ分布を用いて各データ点をランダムに各クラスに割り振り直して算出した 2 指標を比較した．これらの間に有意な差がある場合に，データにはランダムな場合と比較してクラスタが形成されている傾向があるいえる．

この分析の結果，低次のクラスは layer1 からランダムな場合と比較して有意に小さな標準偏差を持つことがわかった．一方，高次のクラスにおいては layer5 のみでランダムな場合との間に有意な差をもつとわかった．一方，クラスタ間平均距離は，低次のクラスでも高次のクラスでも全ての層でランダムな場合と比較して，有意に離れたクラスタが形成されることが確認された．また高次のクラスタにおいて holm 法を用いた多重比較を行ったところ，クラスタ間平均距離に関して layer5 と layer2,3 の間に優位な差がみられた．このことは，layer5 でのクラスタ間平均距離がその他の層に比べて大きい傾向があることを示唆する．

これらの結果から，低次のクラスでは layer1 からクラス毎にクラスタが形成される傾向がある一方で，高次のクラスでは layer5 になって初めてクラスタが形成される傾向があるということが言える．これは，データセット分布の幾何構造とその意味的階層構造が関係することを示唆する．ただし，これらの結果を確定するには，より多くのサンプル数での検証や，より深い層や違うハイパーパラメータで学習された DNN での検証が必要であると考えられる．

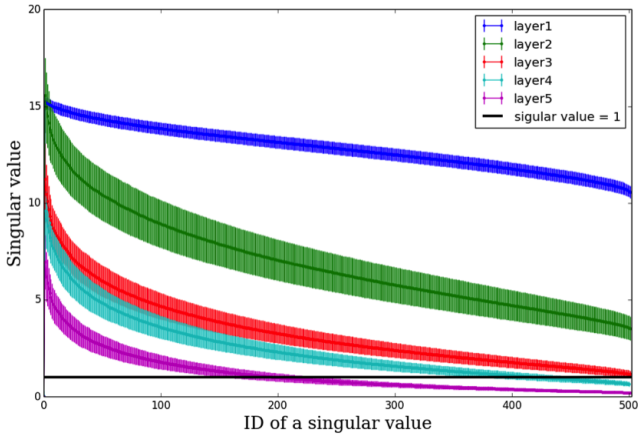

図 4.4: 特異値分布（ImageNet データセット）
AlexNet の特異値分布．層をへる毎に分布が急峻になり，第 5 層では，1 よりも大きな特異値が 200 未満にまで減少している．入力空間が 15 万次元程度あるので，これは非常に小さな次元である．エラーバーは標準偏差をあらわす．

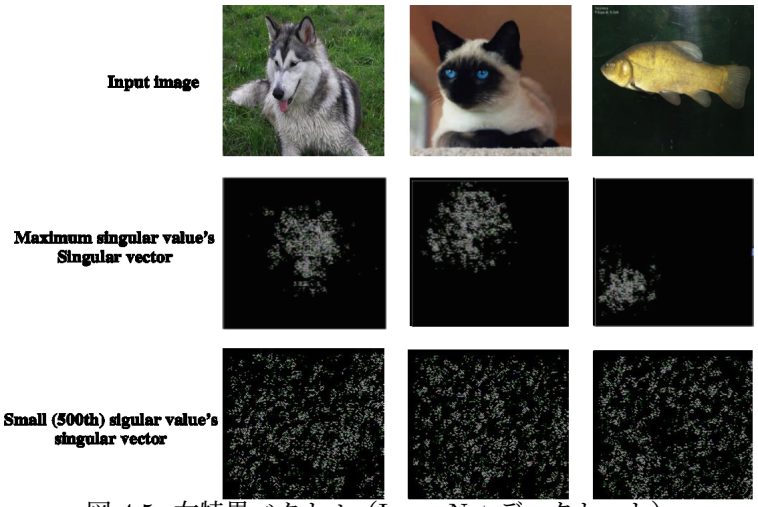

図 4.5: 右特異ベクトル（ImageNet データセット）
大きな特異値に対応する特異ベクトルは，画像の認識対象物の位置や形状に対応したベクトルとなっており，多様体の接線ベクトルとなっていると考えられる．

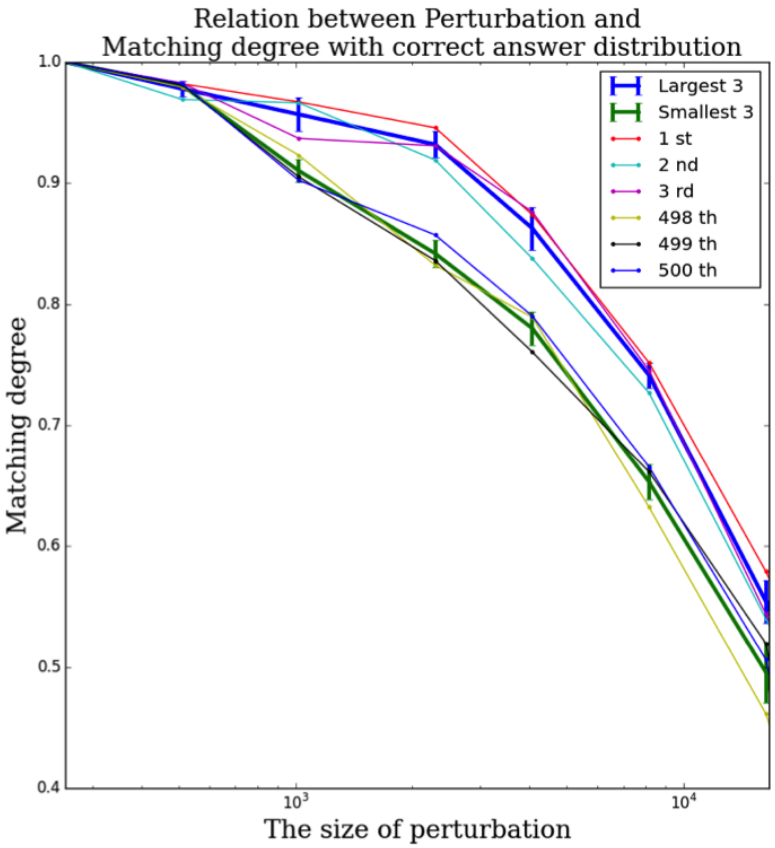

図 4.6: 入力ベクトルの摂動に対する出力のロバスト性
多様体の接線方向に摂動を加えた場合の出力値の元の出力値との一致率（青線）と，垂直方向へ摂動を加えた場合の一致率（赤線）．水平方向の摂動に対して，ネットワークの出力はよりロバストであり，これは多様体の存在を示唆する．

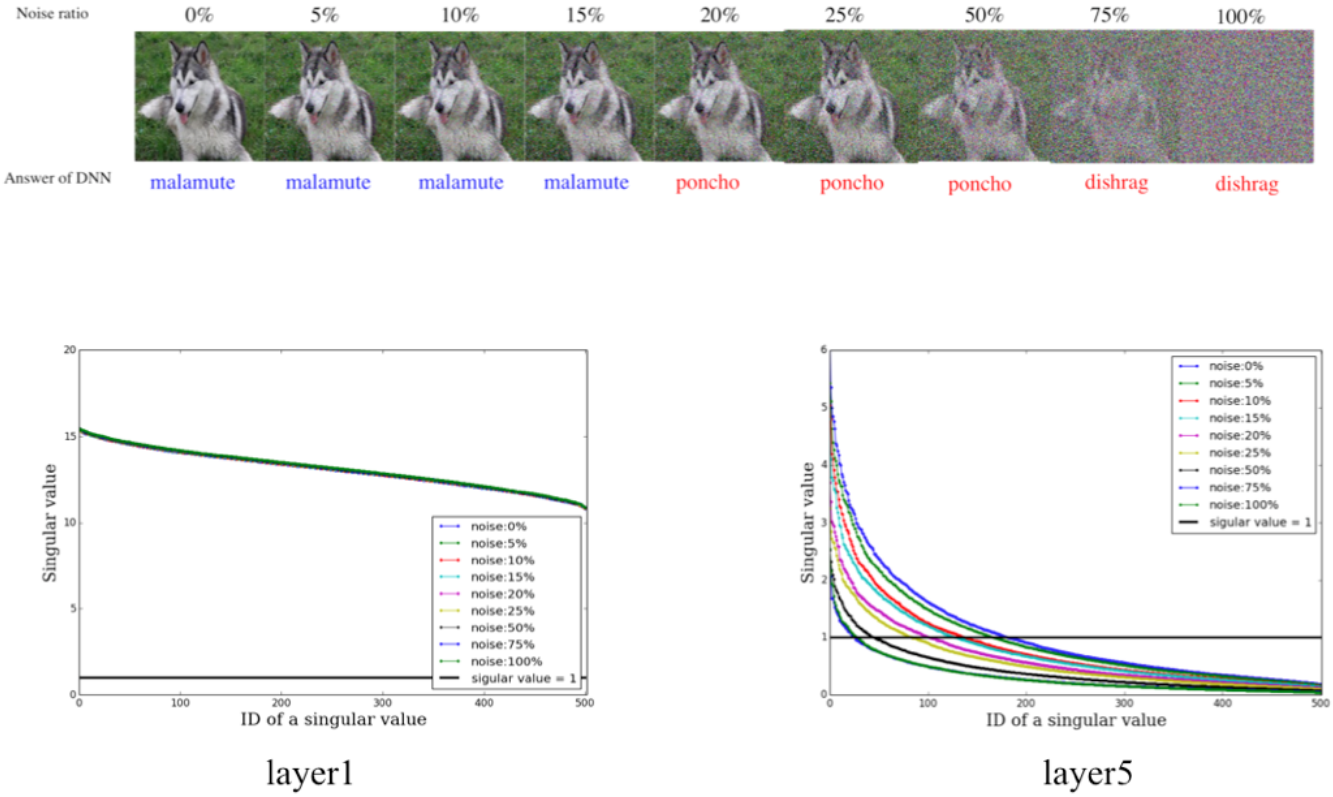

図 4.7: 摂動と特異値分布の関係
上段：ピクセルをノイズに置き換えていった場合の入力データの例．パーセントは，ピクセルの置き換え率．画像下のラベルはネットワークの出力結果．下段：ピクセルをノイズに置き換えていった場合の layer1 と layer5 の特異値分布の変化．layer5 では摂動を大きくするにつれて，特異値分布全体が小さくなることがみられ，情報の伝播が阻害されていることがわかる．

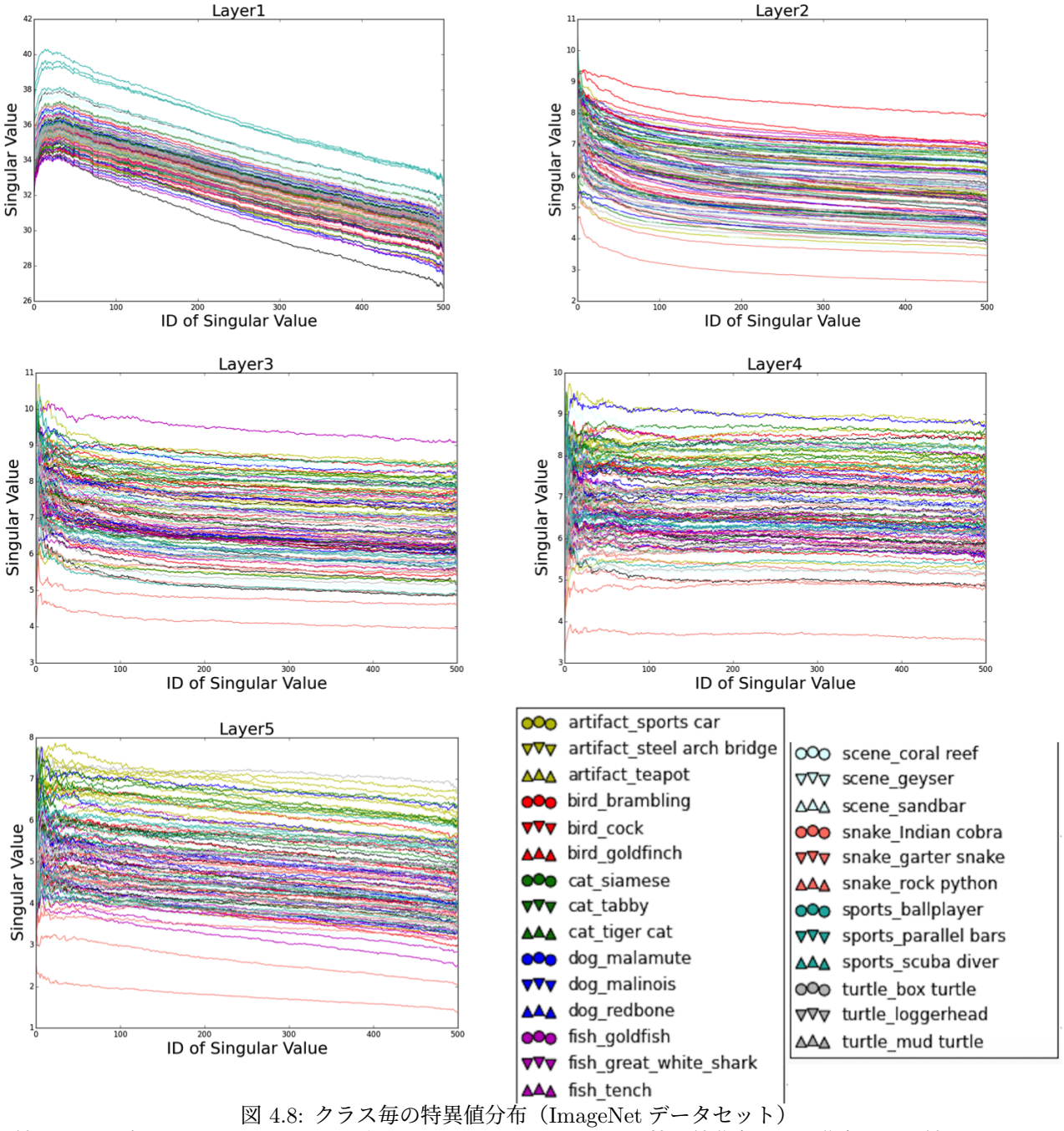

図 4.8: クラス毎の特異値分布（ImageNet データセット）

MNIST の結果ほど明確ではないものの，層を通過するにつれてクラスごとに特異値分布が同じ分布の形に縮退していくようにみえる．

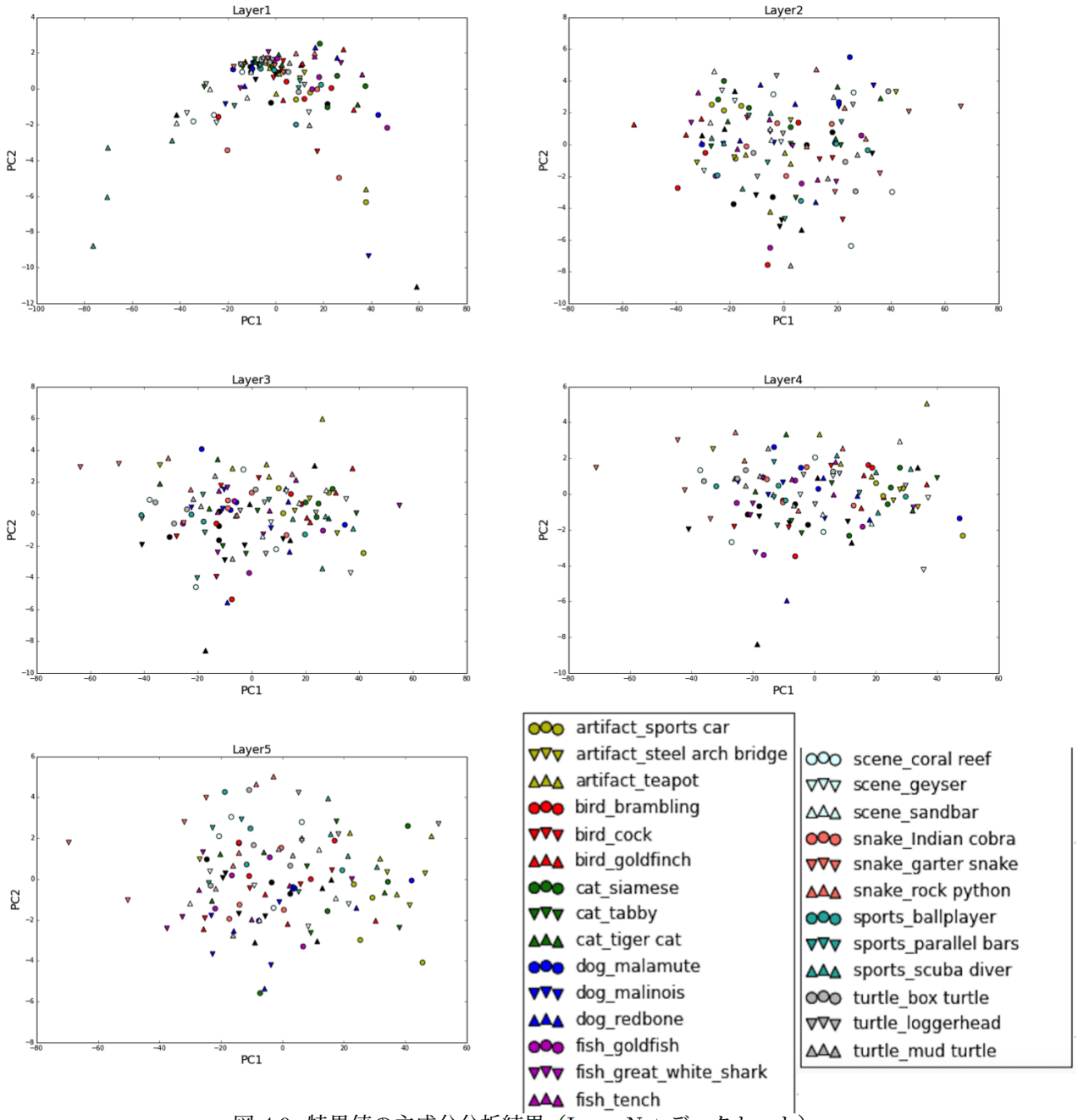

図 4.9: 特異値の主成分分析結果（ImageNet データセット）

層をへるにつれてカテゴリ毎にクラスタを作る傾向がみられた．特に，”siverian husky”は他と比較して，より低次（layer2）の層でクラスタを形成しているようにみえる．一方で，fish や dog 等は，layer5 になってやっとクラスタを形成できているようにみえる．

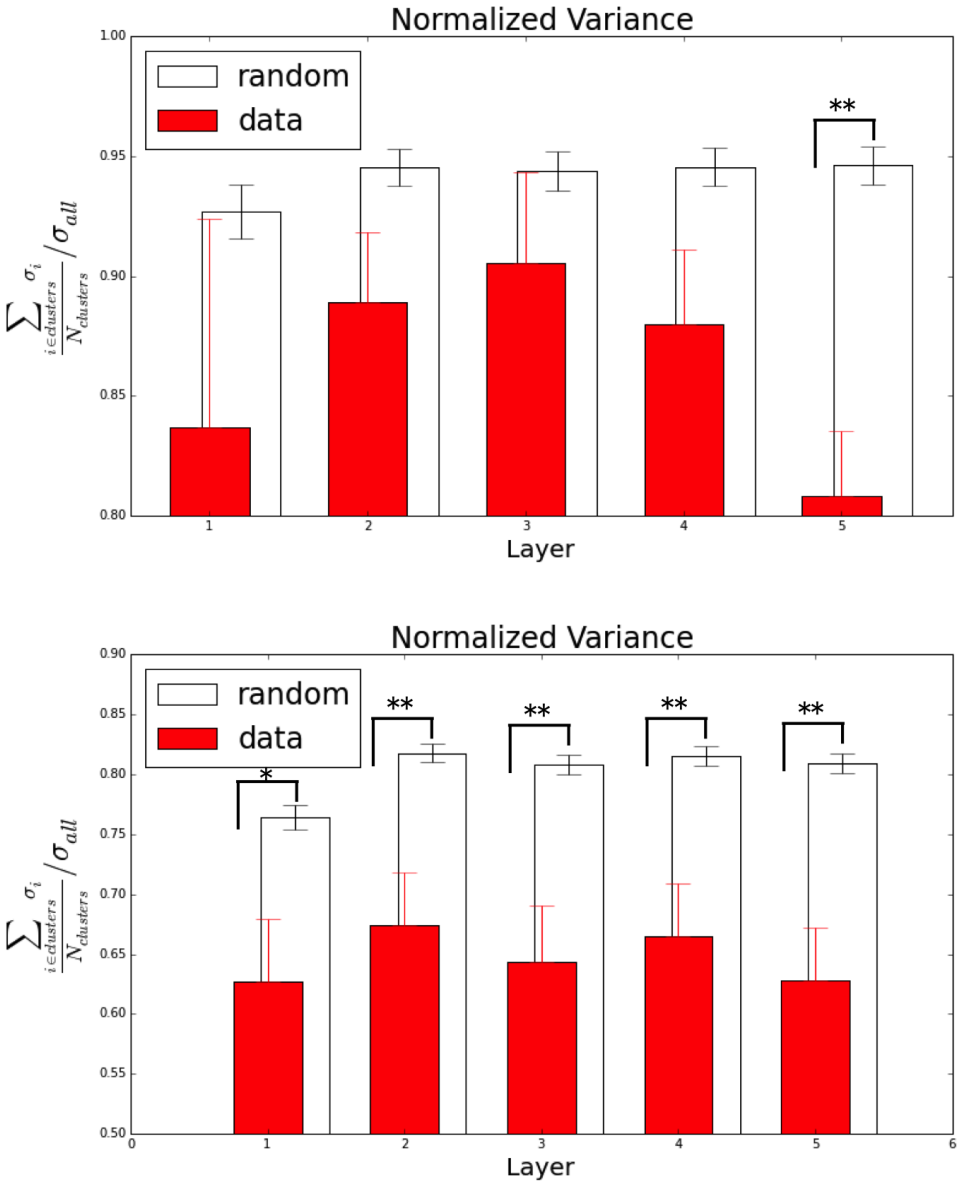

図 4.10: 内的クラスタリング指標：標準偏差
層をのぼる毎にクラス毎の分布がクラスタとしてどのように振る舞うかをみるために，データ全体の標準偏差で規格化されたクラス毎の標準偏差をみた．データからランダムにクラスタを生成した場合と比較して，低次のクラスは layer1 から優位に小さな標準偏差を持つことがわかった．一方，高次のクラスにおいては，layer5 のみで優位に小さな標準偏差をもつとわかった．(正規性の検定を通らなかったため，対応のないノンパラメットリック検定法：Mann-Whitney によって検定した．$*: p < .5, **: p < .1$）エラーバーを標準誤差を意味する．

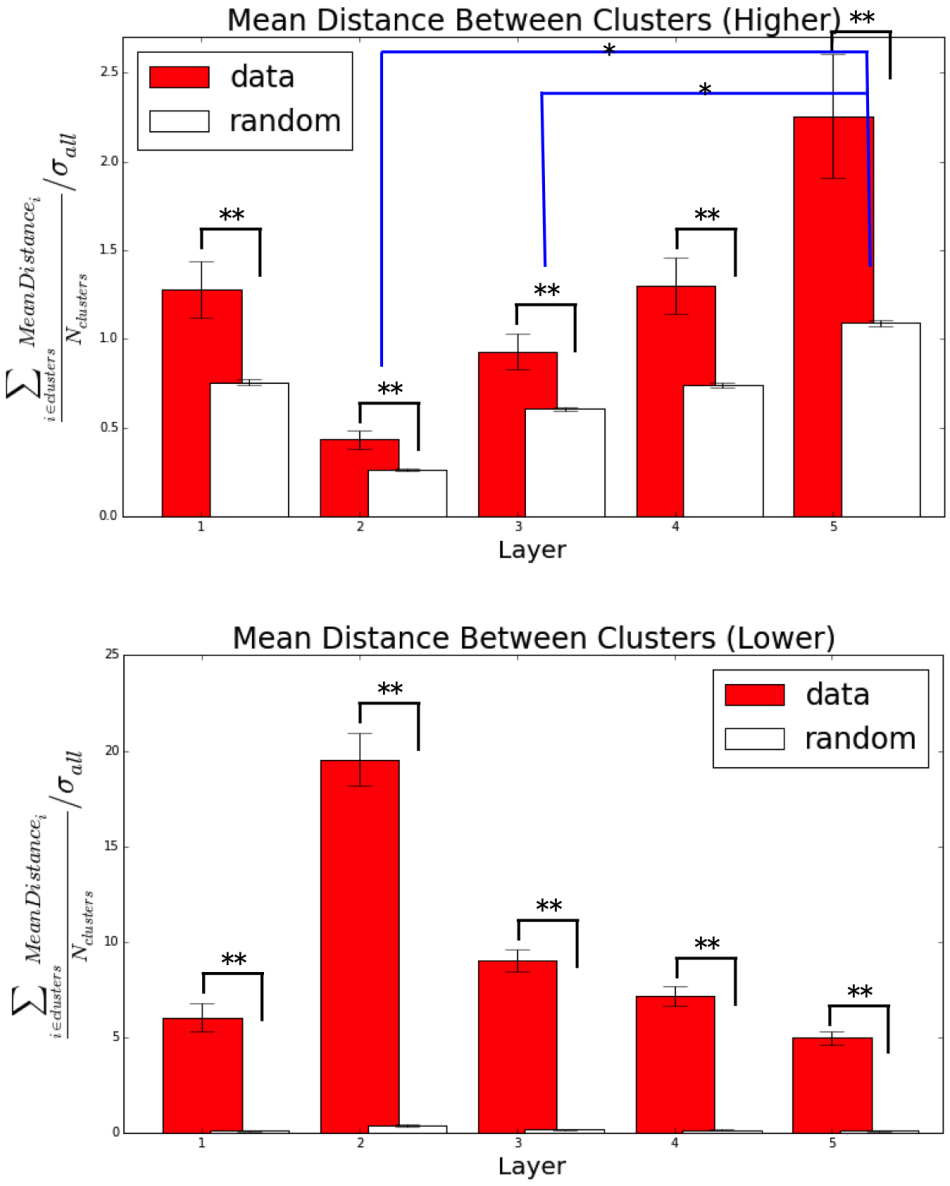

図 4.11: 内的クラスタリング指標：平均クラス間距離

層をのぼる毎にクラス毎の分布がクラスタとしてどのように振る舞うかをみるために，データ全体の標準偏差で規格化されたクラス間の平均距離をみた．データからランダムにクラスタを生成した場合と比較して，低次のクラスでも高次のクラスでも優位に離れたクラスタが形成されることが観測された．特に高次のクラスタにおいて holm 法を用いて多重比較を行ったところ，layer5 と layer2,3 の間で優位な差がみられた．(正規性の検定を通らなかったため，対応のないノンパラメットリック検定法：Mann-Whitney によって検定した．$*: p < .5, **: p < .1$)，エラーバーを標準誤差を意味する．

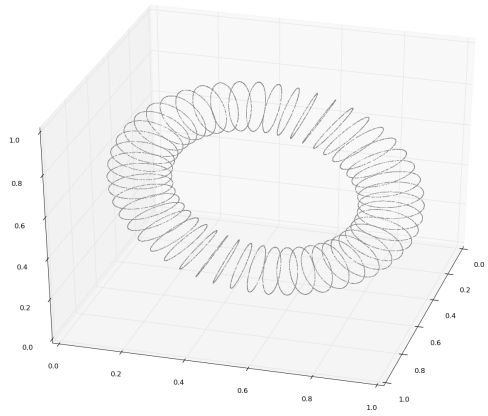

図 4.12: 階層的な幾何構造を持ったデータセットのイメージ
1 つ 1 つの円を低次の構造，それを組み合わせたトーラスを高次の構造と考える．

4.2 幾何的階層構造を持ったデータセットとDNN

4.2.1 目的

前節の最後に得られた，データセットの幾何構造と意味的な階層構造が関係するという結果は，どのように理解すればよいだろうか．本研究ではこの疑問に対して,「データセットの意味的な階層構造がデータセットの幾何構造が対応する」という仮説を設定したい．仮説の具体的な例は図 4.12 のようになる．図のように，小さな円（秋田犬等の低次のクラスに対応）が集まり，大域的なトーラス（犬等の高次のクラスに対応）を形 成するという構造ががあり，これがカテゴリの意味的な階層に対応するというものである．

この仮説の検討のために，階層的な幾何構造を持ったデータセットを学習した際の DNN のふるまいを調べた．具体的には，DNN に階層的な幾何構造を持つデータセットを学習させ，DNN がその多様体を大域的な座標系に写像する機能を獲得できるか確認した．また，階層的な幾何構造を持った多様体が，層をのぼるにつれてどのように写像されていくかという点にも注目して分析を行った．

4.2.2 実験方法

本実験では，大きな円周上に小さな 2 次元多様体（3 次元半球）が並ぶデータセットを用意した（図 4.13）．このデータセットで [3-20-10-2-10-20-3]（図 4.14）という構造を持った Deep Auto Encoder をトレーニングし，前章の人工データを用いた実

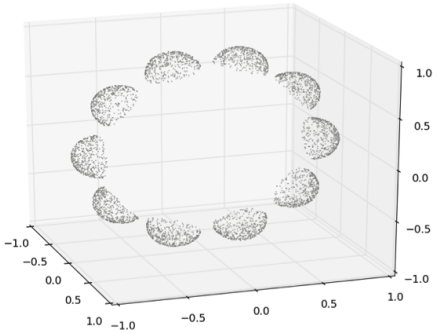

図 4.13: 階層的幾何構造をもったデータセットの分布
階層的な幾何構造をもったデータセット．円周上に半球が並んでいる 2 次元の多様体.

験と同様にしてその DNN を分析した．ここで，説明の便宜のため，DNN の各層を [input,layer1,layer2,layer3,layer4,layer5,layer6] と定義する．また，中間層（layer3）を境にその前層を Encoder その後層を Decoder と呼称する（付録 A 参照）.

さらに，多様体が各階層でどのように変換されるかを調べるために次のような分析も行った．一般に DNN は，シグモイド関数などの活性化関数によって写像関数が非線形化され，空間の場所毎に違う変換（多様体の折り曲げ）を行う．もしシグモイド関数がなければ，写像は単なる線形変換となる．つまり，空間の場所によらず等しく変換が行われるようになる．従って，ある層の活性化関数を

$$f(x) = x \tag{4.2}$$

と線形化すると，その層で行われていた多様体の折り曲げのおおよその様相をみることができると考えられる．そこで，活性化関数を 1 層目から順々に線形化しながら（実際には Decoder 層である layer3→layer6 の対応する層を線形化する.），layer3 をグリッド状に刺激した場合に入力空間で再構成される分布の変化をみて，各層がおおよそどのような写像を担当しているかを分析した.

4.2.3 実験結果

DNN の中間層（layer3）をグリッド状に刺激した結果，実験 1 と同様にニューラルネットが多様体を大域的な座標系に写像しているようにみえることが確認された（図 4.15 の左上図）.

また，1 層目から順に活性化関数を線形化しながら中間層の入力層でのマッピングを可視化した結果，多様体が段々と展開されていく様相が確認された（図 4.15）.

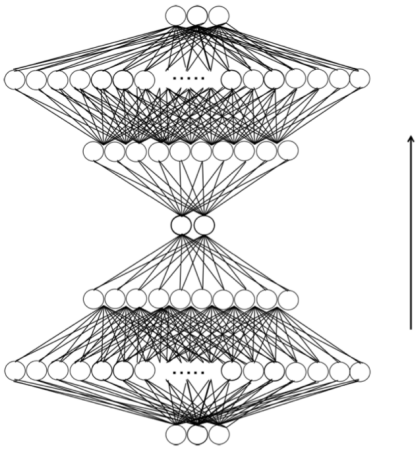

図 4.14: 使用した Deep Auto Encoder

特に，小さな半球構造の展開が layer1 と layer2 の間の写像で行われていること，及び，大域的な円周構造の展開が layer2 と layer3 の間で行われていることがわかった．この結果は，階層的な幾何構造をもった多様体の展開が，折り紙を展開するように，幾何構造の階層性に基づいて段階的に行われることを示唆する．

しかし，ここで行ったのは Decoder 層の可視化であり Encoder 層を直接見たわけではない．Auto Encoder では Encoder 層の重みと Decoder 層の重みが似たものになることが示唆されている [64] ものの，実際に同じような変換を行っているかは自明ではない．従ってここで，Encoder 層の写像関数ヤコビアンの分析によってこれを確認する．

Decoder 層の可視化で見られたようにデータセットの幾何構造が段階的に展開されているならば，2 層目にて半球構造が展開されてクラス（半球）毎に大域的座標系が実現されているはずである．それをみるため，図 4.16 にあるように，3 つのベクトルが同じ 2 次元平面内にある場合，それらのなす角の和は 90 °となるという事実を用いて，各クラス毎の平均左ベクトルを算出した上で，それと各左ベクトルとのなす角を求めた（図 4.17）．その結果，Decoder の可視化によって半球構造の展開が確認された layer2 では，layer1 と比較してなす角が 90 °に近づく傾向があることが確認された．しかし，その傾向は layer3 でも継続してみられる．これは，Decoder の可視化で見られた段階的な構造の展開を否定するものではないが，これだけでは layer2 で各クラスに対応する幾何構造の展開が行われているとはい

い難い．

結果のさらなる検証のために，右特異ベクトルの分析によってDNNが入力空間から多様体構造を抽出できているかを確認した．具体的には，小さな半球の1つ1つに別の基準ベクトルを設けることで，前章と同様の手続きで（図3.7参照）DNNのヤコビアンの右特異ベクトルが小さな半球の接空間を捉えているかを検証した．その結果，layer2において右特異ベクトルの垂直ベクトルが，半球の頂点付近（90°）においてデータから算出された垂直ベクトルの方向とおおよそ一致する傾向があることが確認された（図4.18）．しかし，それは前章の結果と比較して非常に荒い結果であった．前章で議論したように，正方形で多様体を被覆する際に伸縮を必要とする場合，右特異ベクトルで構成される空間と多様体の接空間との間に誤差が生じる．今回のように複雑な図形の場合，各部で伸縮が生じると考えられ，その結果として特異ベクトルの分析では明確な結果が得られなかった可能性が考えられる．また，この誤差は左特異ベクトルにも影響を与えると考えられ，左特異ベクトルの結果の信頼性も低いものと考えられる．

このように，特異ベクトルによる定量的な分析によって仮説は検証できなかった．これを検証するには，多様体の伸縮に影響を受けない分析手法の開発が必要となると考えられる．

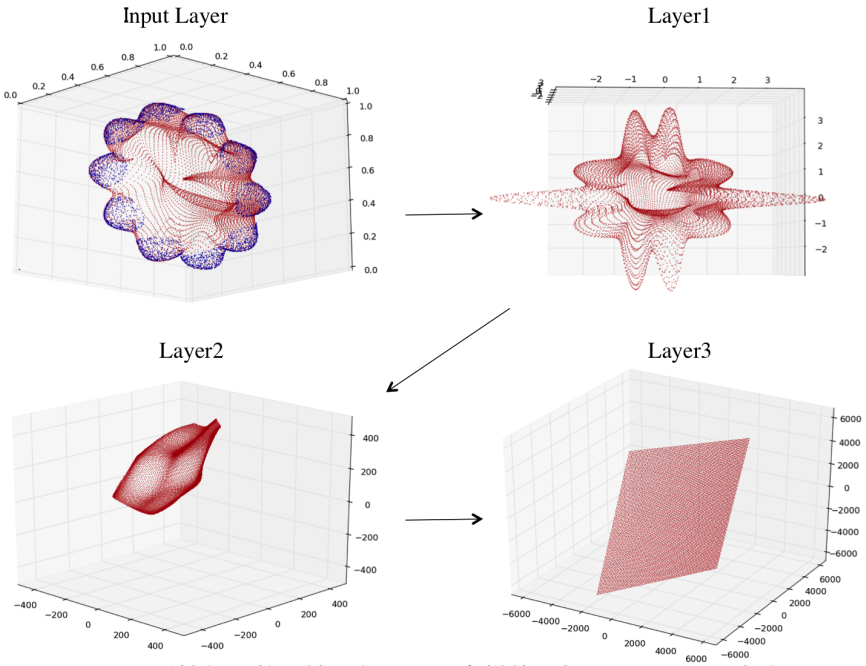

図 4.15: 活性化関数の線形化による多様体の折りたたみの可視化
第 1 層から順次活性化関数を線形化していくことで，各層でおおよそどのような多様体の折りたたみを行っているかを可視化している．

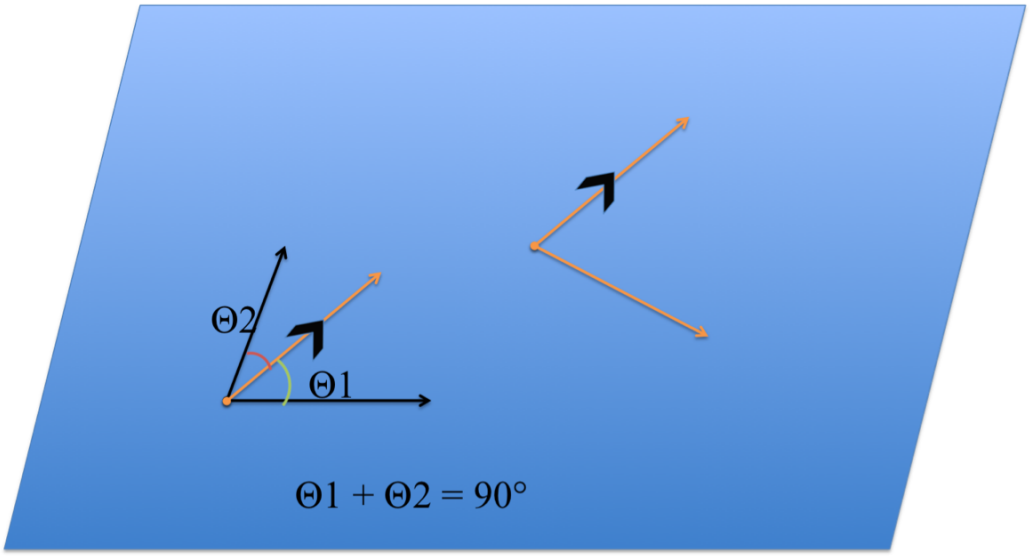

図 4.16: 同一平面条件
もし 3 つのベクトルが同一平面内にある場合，それらのなす角の和は 90 °となる．これを同一平面にある指標として分析を行った．

4.3 まとめと考察

本章では，教師あり学習アルゴリズムによってトレーニングされた DNN（Convolution Neural Networks）を用いて，意味的な階層構造を持った大規模データセットである ImageNet データセットの観察を試みた．分析の結果，特異値分布が急峻になること，及び右特異ベクトルが MNIST の観測時と同様に，1 より大きな特異値に対応するベクトルは入力画像に対応するような局所的な構造を持ち，1 より小さい特異値い対応するベクトルは広範囲に分布するノイズのような構造を持つことが確認された．これらの結果を踏まえ，より明確に ImageNet データセットが多様体構造を持つことを確認するために，算出した右特異ベクトルを用いて，入力画像に多様体の接線方向と垂直方向に別々に摂動を加えた際の DNN の挙動を検証した．この結果，水平方向の摂動は垂直方向と比較して出力の変動がロバストであることが確認された．さらに，垂直方向の摂動の影響も吸収されるような吸引領域も確認された．これらの結果は，ImageNet データセットが多様体仮説を満たすような構造を持つことを意味するとともに，観測に利用した教師ありでトレーニングされた Convolution Neural Networks が，入力空間から多様体構造を抽出する能力を持つことも示唆する．

ImageNet データセットのような大規模な画像データセットが多様体構造を持つ

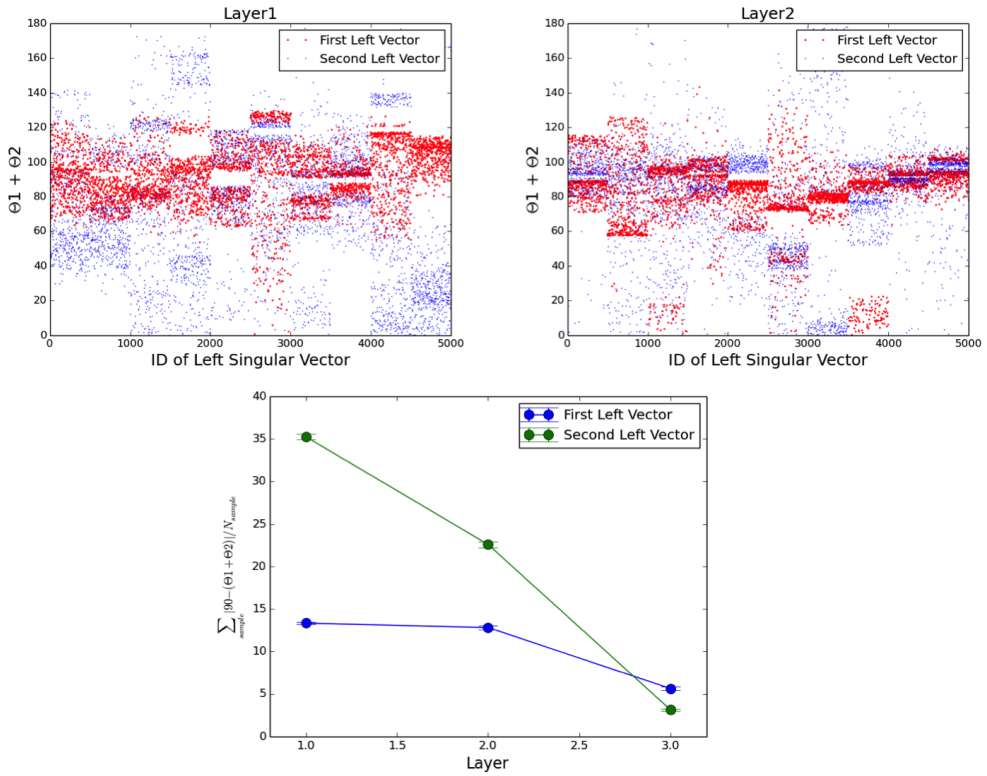

図 4.17: 左特異ベクトルによるクラス毎の幾何構造展開状況の検証
上図：左特異ベクトルの第 1，第 2 成分と，それらのベクトルのクラス毎平均ベクトルとの偏角の分布．　下図：上図を元に計算した平均偏角．縦軸は偏角と 90 °との差分値である．従って，0 になるほどより大域的な座標系となっていることを意味する．エラーバーは標準誤差．N=5000

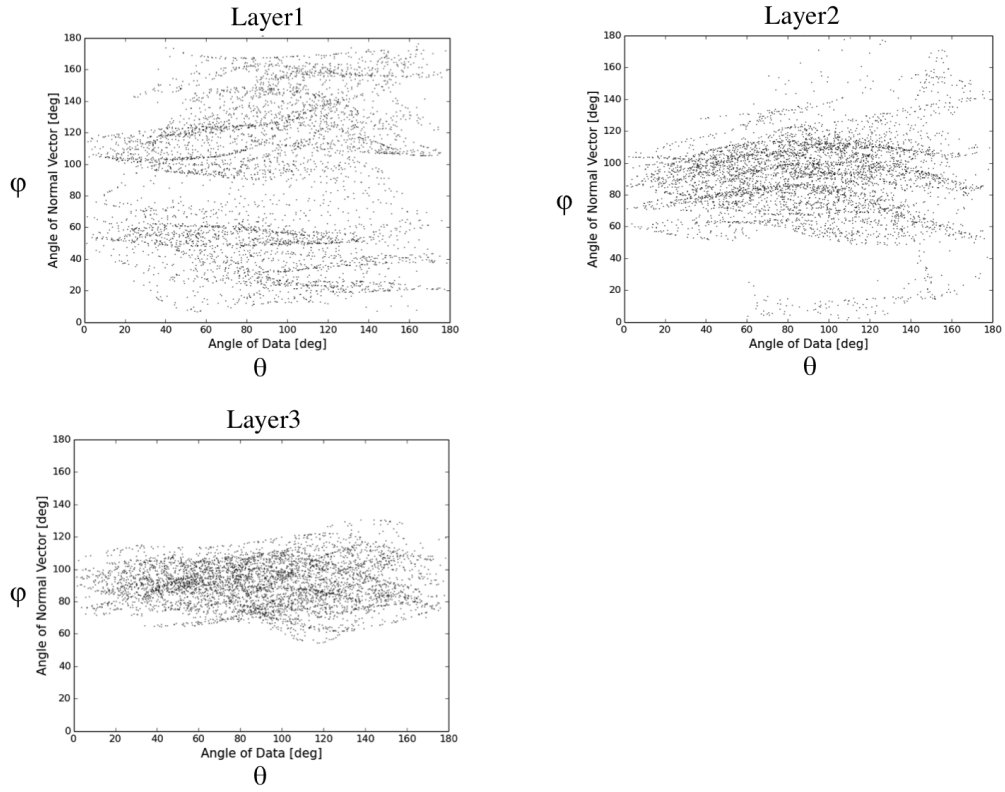

図 4.18: 右特異ベクトルと接空間の垂直ベクトル（階層的幾何構造）
各球に対して，先の偏角（図 3.7）を定義し，右特異ベクトルの垂直ベクトルと多様体の垂直ベクトルを比較した．かなり粗いものの，layer3 において 90 °付近に θ と φ が収束することが確認される．これは，DNN が半球の頂点付近を捉えていることを示唆する．

ことを定量的に示した研究はこれまでにない．また，教師ありでトレーニングされた Convolution Neural Networks が多様体を抽出する機能を持つことを示す結果が得られたのも，本研究が初めてであると考えられる．

さらに，MNIST の場合と同様に多様体の次元や曲がり具合といった情報が含まれる特異値分布とデータセットの意味的なカテゴリが関係しないかを検討した．ImageNet データセットの意味的な階層構造に対応して高次のクラスと低次のクラスを定義した上で，それらのクラスに対応するクラスタの性質を調べた．ランダムに設定したクラスのクラスタとの比較から，低次のクラスは layer1 の時点でクラスタが形成される傾向があることが，高次のクラスは layer5 になって初めてクラスタが形成される傾向があることが判明した．このことは，データの意味的な階層構造がデータの幾何構造と関係することを意味する．このような関係性はこれまでには報告されていない．またこれが正しい場合，特に文化的要因や学習によらない人間の低次の表象の形成において，データの何らかの幾何学構造が重要な意味を持つことになり，その場合，本研究の結果が認知科学等の領域に与える影響は大きいものと考えられる．

本研究ではさらに，データセットの意味的な階層構造と幾何構造を対応づける仮説として,「データセットはデータの意味的階層に対応じた幾何学的な階層構造を持つ」という仮説を提案し，その妥当性について検討した．具体的には階層的な幾何構造を持ったデータセットでトレーニングされた DNN を分析した．その結果，DNN が階層的な幾何構造を持ったデータセットを大域的な座標系へ写像する機能を獲得できることがわかった．さらに，Decoder 層の写像機能の可視化から，DNN が多様体を折り紙を展開するように段階的に変換しながら，大域的な座標系へ写像していく描像が得られた．しかしながら，Encoder 層のヤコビアンの分析からはこれを確認することはできなかった．今後のさらなる検討が必要と考える．

ところで，DNN が画像などのデータセットを写像する際，階層をのぼる毎により抽象的な表象へと段階的に変換されることが示唆されている [40][39]．この事実は，DNN が階層的な幾何構造を持つ多様体を折り紙を展開するように段階的に展開していくという描像を支持すると考える．なぜならば，抽象化のためには抽象化する領域を n 対 m（m<<n）写像する必要があり，その場合，複雑な領域の検出と抽象化を同時に行うよりも，抽象化する領域を先に大域的座標系へ展開しておき，その後抽象化した方が効率が良いと考えられるためである．ここでは，段階的に展開される幾何構造が意味的なクラスに対応すると考えている．

第5章　大自由度力学系のもつ多様体構造と深層学習

5.1　大自由度力学系と多様体

5.1.1　データの種類と多様体仮説

これまでに見てきた画像データ以外に，どのような種類のデータが多様体構造を持つだろうか.

例えば音声・音楽データのパターン認識において，多様体を抽出するように正則化項を導入したアルゴリズム [57][58] や対称性に基づくフィルタをかけること [68] が有用であることが示唆されており，それらのデータが多様体構造をもっていることが強く示唆されている.

また，大自由度力学系の分野では，多様体仮説とは別の視点から，データセットの多様体構造の抽出によるそのデータの縮約表現の取得が議論されている．例えば，タンパク質のフォー ルディング現象の時系列データからISOMAP によって多様体構造を抽出したところ，2 次元程度の次元で自由エネルギー地形のゆらぎの 9 割以上を捉えることに成功したという報告がある [19].

このような大自由度力学系の 1 つに，動物などの集団としての複雑な運動をモデル化したボイドモデルを大自由度化したものがある．後述するように，これまでの集団運動の研究では，大自由度ボイドモデルのシミュレーションで観測されるような複数種類の「群れ」が相互作用するような複雑な現象は，活発には研究されてこなかった．この要因の一つとしては，そのような複雑な条件下で，性質の違う「群れ」を抽出することが困難であったことがあると考えられる．そこで本研究では,「群れ」を「ある同一の多様体に，その時系列データが埋め込まれた個体の集合」と定義した上で，これを抽出することを試みた．ただし，ここで1つの問題が生じる．前述したような既存の多様体抽出アルゴリズムは，写像関数を学習するようなアルゴリズムにはなっておらず，全てのデータを同時にメモリ上に展開した上で，それを一気に変換するアルゴリズムとなっている．この条件は，特にコンピュータのメモリの制約から，大自由度の力学系の分析を行うことを困

難にする [15]．一方で，DNN は写像関数を学習することができ，学習もデータの一部ずつを逐次メモリに展開しながら行うことができるため，大自由度の力学系の分析に適していると考えられる．そこで，本研究でここまでに得られてきた知見である，パフォーマンスの高い DNN は多様体構造を抽出できるという知見に基づき，DNN を用いて「群れ」に対応する多様体構造を抽出することを試みることを考えたい．

一方で，DNN の学習においてハイパーパラメータのチューニングは非常に困難な課題である [8]．そこで本研究では，これまでに得られた知見に基づき，特異値分布をより急峻にすることを指標としてハイパーパラメータチューニングすることを試みた．そのためにまず，人工データを用いた検証実験を行い，その上で実際の「群れ」抽出学習にこれを適用した．

5.1.2 本省のながれ

以上のような背景に基づき，本章では次のように議論を展開する．

まず，対象とするボイドモデルについて解説する．その上で DNN による分析との比較のため，まず大自由度ボイドモデルのシミュレーション結果を既存手法で分析し，その結果を解説する．次に，DNN の適用の前提となる 2 つの点を確認する．1 つ目は，多様体構造の抽出によって多様な種類の「群れ」を同時に抽出できるかの確認で，少数自由度のボイドモデルの時系列データと t-SNE 多様体学習アルゴリズムを用いて検証する．2 つ目は，特異値分布を DNN のハイパーパラメータチューニングの目安とできるかの確認で，人工データと Deep Auto Encoder を用いて検証する．そして最後に，DNN による「群れ」構造の抽出を行い，得られた結果を既存手法の結果と比較する．

5.2 力学系：ボイドモデル

5.2.1 群れ運動

我々のまわりには，動物たちによる数多くの集団運動が観察される．敵の存在によって魚達はトーラス状の群れを形成し，数百匹のムクドリによって構成される群れは，それがまるで 1 つの個体であるようにまとまって動く [61]．また，我々人間も，スクランブル交差点でみられるように，集団の流れのダイナミクスを持つ [61]．このようなマクロなスケールだけでなく，細胞やたんぱく質といったナノスケールの世界でも多様な集団運動が観察される [61] [18] [49] [10] [42].

さらに，cavagna らは，ムクドリの鳥の群れの実測とその分析を通して，その群れの個体の速度ゆらぎの相関長と群れのサイズの比が，群れのサイズによらず一定であるというスケールフリー性の存在を発見した [11]．この結果は，集団が 1 つの個体として振舞うようなシステムが，臨界システムであることを示唆しいる [11]．また，個体間の単純なルールによってこの集団運動を再現する，ボイドモデルと呼ばれる一連のモデルが提案されている ([2][44][46][32][60] [16])．例えば Vcsek は，鳥の群れの一個体の速度ベクトルを強磁性体におけるスピンになぞらえ，集団運動の次のようなモデル化を行った [60]．モデルは次式で定義される．

$$\theta_i(t+1) =< \theta_j(t) >_r +\eta_i(t) \tag{5.1}$$

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + v_0(cos\theta_i(t+1), sin\theta_i(t+1)) \tag{5.2}$$

$$\tag{5.3}$$

ここで，$\theta_i(t)$ は時刻 t での個体 i の運動方向を表し，$< \theta_j(t) >_r$ は個体 i を中心にした半径 r 以内にいる全個体の平均運動方向を表す．また，$\eta_i(t)$ は，$[-\eta/2, \eta/2]$ の一様分布のノイズ項であり，各個体は一定の速さ v_0 で移動する．つまりこのモデルは，一定のノイズの元まわりと運動方向を揃えながら運動を行う集団をあらわす．このノイズ項の大きさを変えることで，系全体が秩序層（集団運動層）から不秩序層へと変化するような現象が生じる．このモデルでは，集団運動層での局所的な粒子密度揺らぎが，中心極限定理で期待される値より大きくなる巨大密度揺らぎの存在が確認されており，非平衡系特有の様々な特徴的等計測を有することから多くの研究が行われきた [69]．

一方で，この Vicsek モデルでは，実際の群れにあるような明確なクラスタのようなものは形成されにくい．実際 chate らは，Vicsek モデルのような方向を揃える効果以外に，個体同士が近づこうとする引力と，近づきすぎるとぶつからないように離れようとする斥力の効果を加えることで，3 次元でのシミュレーションにおいて，より自然の鳥の群れに近い動きが生成されることを報告している [14]．例えば Couzin らは，このような 3 つの力によるボイドモデルで生成される群のパターンとして，群れの平均速度や平均角運動量で区別される，図 5.2 にあるような 4 種類の群れが生成されることを報告している [16]．

5.2.2 群れの相互作用

このように，ダイナミックな秩序構造を作り出すモデルとしてボイドモデルの研究が多く行われている．一方で，それらの研究は，1,000 匹オーダー程度の個体

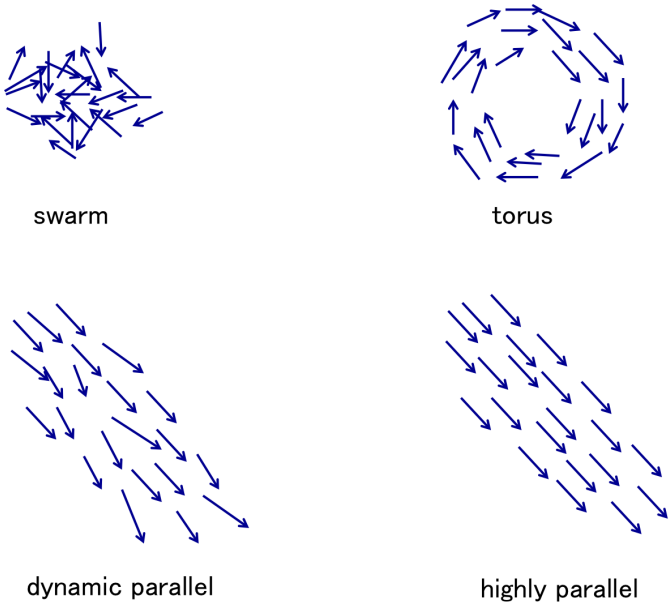

図 5.1: ボイドモデルにおける群れの例
左上から時計回りに，虫の群れのように個体がランダムに動きながらも固まりとなる swarm（平均速度：低，角運動量：低），魚の群れにみられるトーラス（平均速度：低，角運動量：高），ムクドリの群れのように多様な振る舞いを示す dynamic parallel（平均速度：中，角運動量：低），そして，一方向に一斉に運動する highly parallel である（平均速度：高，角運動量：低）.

数でのシミュレーションで，1つの群れの振る舞いを対象としたものが中心となっており，特に複数の群れが同時に相互作用するような大規模な個体数での研究は少ない．

これに対して，自然界の群れ同士は相互作用しており，例えば西洋ニシンの群れにおいて，栄養状態に応じて群れが，海中の上層にいる群れと下層にいる群れの2つに分離し，さらに，各個体の栄養状態の変化に応じて，それら2つの群れの間でメンバーが入れ替わるようなダイナミクスが観測されている [4]．従って，集団運動に対する新たな知見を得るためには，群れが複数存在するような大規模な系でのシミュレーションとその結果の解析を行うことが必要であると考えられる．以降では，これを行なっていく．

5.3 大自由度ボイドモデルのシミュレーション

5.3.1 Reynolds のボイドモデル

本研究で目指す，群れが複数存在するようなシミュレーションには，明確な形を持った群れを生成する必要があるため，前述したように3つの力を用いるモデルが最適であると考えられる．本研究では，このようなモデルの中から，Reynolds 等によって提案されたボイドモデル [46] を選択した．このモデルは，相互作用から群れ行動を再現しようとする最初期のモデルの一つであるとともに，複雑な鳥の群れの動きを再現できることがわかっている．このような実績を考慮し，このモデルを選択した．

前述したように，Reynolds のボイドモデルは,「引力」「斥力」及び「まわりと同じ方向へ進もうとする力」の3つの個体間相互作用によって，群れとしての多様な振る舞いを再現する．Reynolds の論文 [46] では，相互作用の定性的な定義のみしており定式化はされていない．そこで，本研究では Reynolds の論文 [46] に基づき，具体的には以下のような定式化の元シミュレーションを行った．

$$\Delta\vec{v}_i = W_{att}\cdot(\vec{x_i} - \frac{\sum_{j\in S_{att}}\vec{x_j}}{n_{att}}) + W_{rep}\cdot(\sum_{j\in S_{rep}}\frac{(\vec{x_i}-\vec{x_j})}{|(\vec{x_i}-\vec{x_j})|}) + W_{ali}\cdot(\vec{v_i} - \frac{\sum_{j\in S_{ali}}\vec{v_j}}{n_{ali}})$$

ここで，x_i は i 番目の粒子の位置をあらわす．また，左辺の第1項が引き寄せ合うルールを，第2項が反発するルールを，第3項が速度を揃えようとするルールを表しており，それぞれのルールはそれぞれの粒子間距離がある範囲内にあるときだ

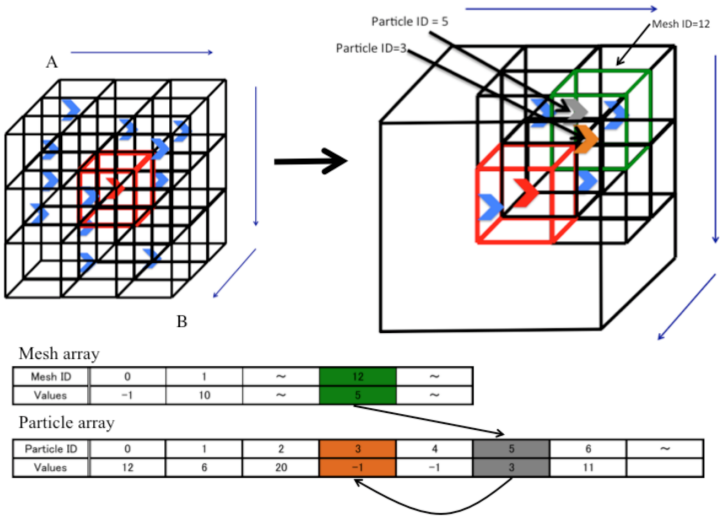

図 5.2: 大規模シミュレーションの方法
メッシュの分割法と，粒子データのメモリ割り当て法.

けで働く.（それぞれ，W_{att}，W_{rep}，W_{ali} がそれをあらわす.）速度の大きさは V_{max} と V_{min} の間で変動する.

5.3.2 大規模シミュレーション

大規模ボイドモデルのシミュレーションのために，本研究ではGPGPU（General-purpose computing on graphics processing units）による並列演算を用いた.

具体的なアルゴリズムは次の通りである．スレッドを粒子数分生成しそれぞれの粒子についての計算を並列化した．この時，空間をメッシュに切り分け，相互作用の及ぶ範囲内のメッシュについてのみ相互作用を計算するようにして，演算量を減じた.

各メッシュには複数の粒子が入りうるが，GPU のメモリサイズの制約から，各メッシュ毎に粒子を複数個登録できるような長さの配列を用意することはできない．従って，da Silva らの方法 [1] に従って，次のような手順で粒子をメッシュに対応付 けた（図 5.2 参照）.

1. mesh array と particle array を用意し-1 で初期化する.
2. あるメッシュに粒子が存在する場合, 対応する mesh array の要素にその粒子の ID を代入する.

3. 1つ目の粒子のIDをmesh arrayに代入し, 次に1つ目の粒子に対応するparticle arrayの要素に, 他の1つの粒子のIDを登録する.

4. 次の粒子についても順次particle arrayに登録していく.

5. 読み取り時は,mesh arrayからIDを順番に参照していき, 値が-1になった場所 で読み取りを終了する.(図下方参照)

この並列化によって，複数の群れが形成されるようなパラメータ領域において，50万匹程度のシミュレーションを1[fps]程度の速度で行うことが可能となった.

計算に用いたGPUはGTX Titan Blackで，このGPUのメモリサイズは6GB, プロセッサ数は2880基，単精度の浮動小数点演算速度は5.12TFLOPSであった.

5.3.3 シミュレーション結果

表5.1のパラメータの元，大規模な個体数でのシミュレーションを行った．シミュレーションの空間は3次元の周期境界空間となっており，各個体は速度の上限と下限をもつ.

このパラメータにおけるシミュレーションの結果，多数の群れが相互作用する状態が現れた.（図5.3）

結果の定性的な観察として，そこには全員が方向を揃えて進む小さなフィラメント状の群れと，群れの内部で個体がランダムに運動しているようにみえる，大きな淀みの群れの2種類が存在することがわかった．また時間発展に従って，大きな群れからフィラメント状の群れが出現することや，逆にフィラメント状の群れが大きな群れに吸収されることなどが観察され，これら種類の違う群れが相互に作用しつつ，全体の状態を形成していることがわかった．また，長時間のシミュレーションによってこの状態が長期間安定的に維持されることも観察された．このような観察結果をもとに以下で定量的な分析を行っていく.

5.4 既存手法での分析

5.4.1 分析結果

シミュレーション空間をメッシュ状に分割し，各メッシュの個体数密度の頻度分布をみると，ベキ分布とガウシアン分布の和となるような分布がみられた（図5.4）．それぞれの分布を抽出し，その分布内の粒子同士の最近傍距離を算出し，粒子間距離に基づくクラスタリングアルゴリズムであるDBSCAN（Density-Based

Parameter	Value	Parameter	Value
range of cohesion	0.05 [unit]	angle of cohesion	$\pi/2$
range of alignment	0.05 [unit]	angle of alignment	$\pi/3$
range of separation	0.01 [unit]	angle of separation	$\pi/2$
field size	0.5 - 3.0 [unit]	number of individuals	2,048 - 524,288
max speed	0.005 [unit/step]	min speed	0.001 [unit/step]
W_{att}	0.008	initial velocity	randomly distributed
W_{rep}	0.002	initial position	randomly distributed
W_{ali}	0.06	time step increment	1
density	16,384 [num/unit^3]	boundary condition	3D periodic

表 5.1: Parameters of this simulation

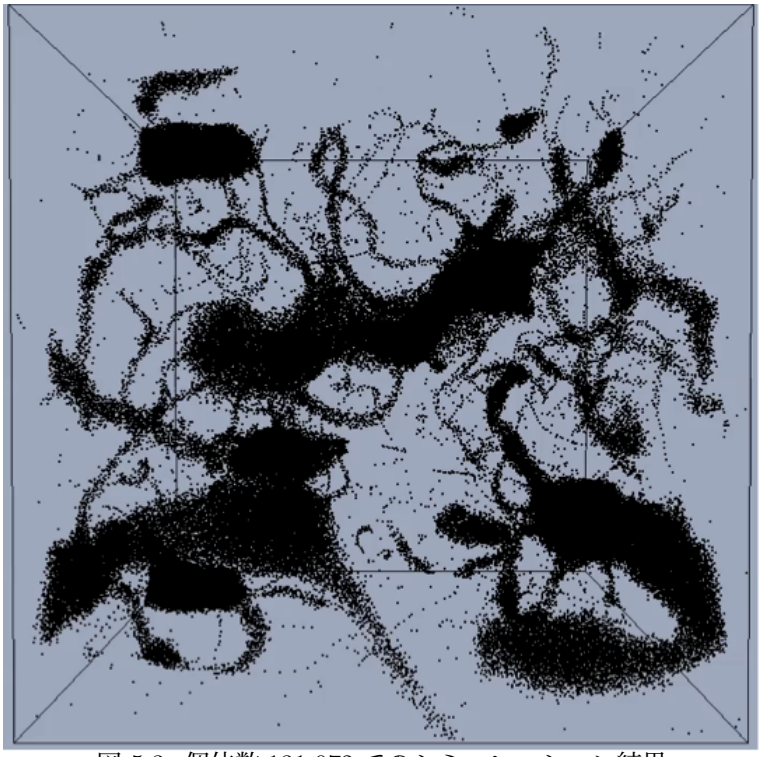

図 5.3: 個体数 131,072 でのシミュレーション結果
大きな塊の群れと，その間を飛び交うフィラメント状の群れの 2 種類の群れがみえる．

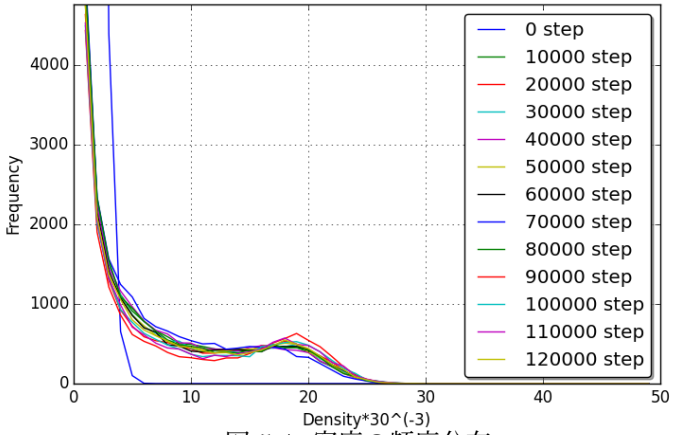

図 5.4: 密度の頻度分布
ベキ分布とガウシアンが重なった分布が見られる．複数の曲線はシミュレーションの step 数を表し，10,000 ステップ以降は，この分布が安定的に存在することがわかる．

Spatial Clustering[23]）を用いて，粒子のクラスタリングを行った．すると，図 5.5 のように，大きな群れ（ガウス分布に対応）と，フィラメント状の群れ（ベキ分布に対応）の抽出に成功した．

さらに，各群れの個体数の分布をプロットすると，10,000 オーダーの群れと，1.000 オーダー以下の小さな群れとの間に，その大きさの群れが存在しないギャップがみられた（図 5.6）．このギャップよりも大きな群れがガウス分布に対応する大きな群れになっており，このギャップの下にはフィラメント状の群れが存在した．

また，群れ毎のエネルギー分布をみると，やはり 10,000 オーダーの大きな群れと小さな群れの間で違いがあり，大きな群れはエネルギーが小さく小さな群れはエネルギーが大きかった（図 5.7）．これは，定性的な観察結果と一致する．ちなみに，大きな群れについては，群れの表面が内部に比べて早い速度で運動していることもわかり（図 5.8），外を運動する個体によって内部の個体が閉じ込められているというダイナミクスの表像が得られた．

このように，シミュレーションのパラメータが同じにもかかわらず，局所的にこのようなメカニズムの違いが生じることが観察されたが，この現象は，もし機械学習によって群れが抽出できなかった場合は，みられなかった現象であるといえよう．

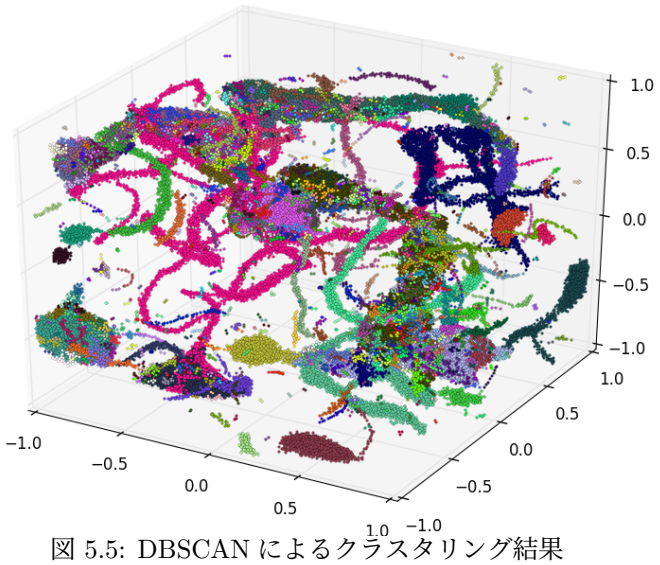

図 5.5: DBSCAN によるクラスタリング結果
DBSCAN による密度分布毎のクラスタリングによって，大きな群れとフィラメント状の群れが抽出できていることがわかる．

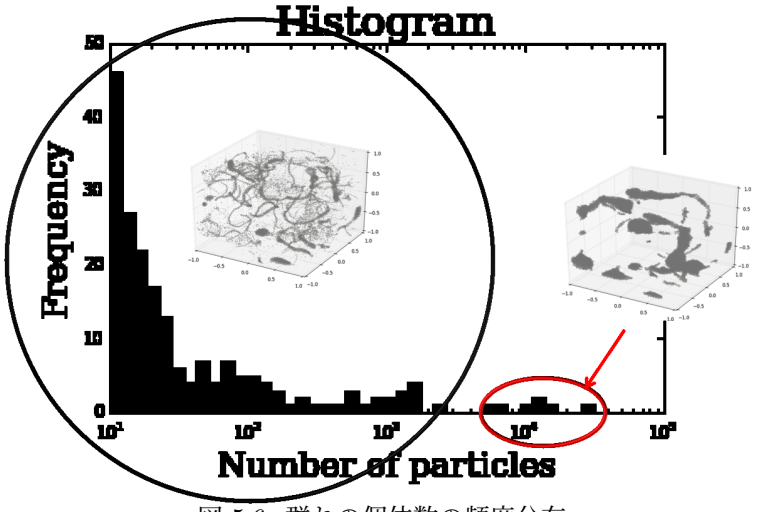

図 5.6: 群れの個体数の頻度分布
3,000 匹となる群れを境に大きな群れとフィラメント状の小さな群れに分離される．

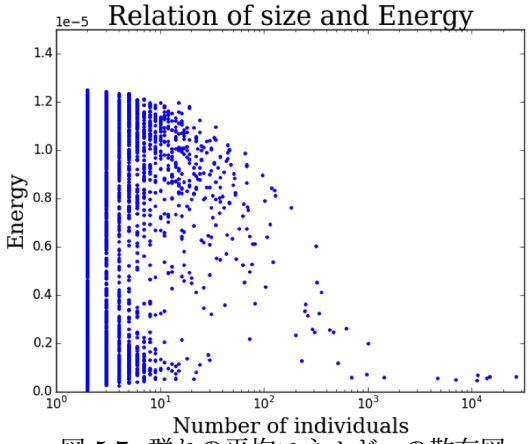

図 5.7: 群れの平均エネルギーの散布図

クラスタリングで得られた群れ毎の平均エネルギーの散布図．大きな群れほどエネルギーが小さく淀んでいることがわかる．

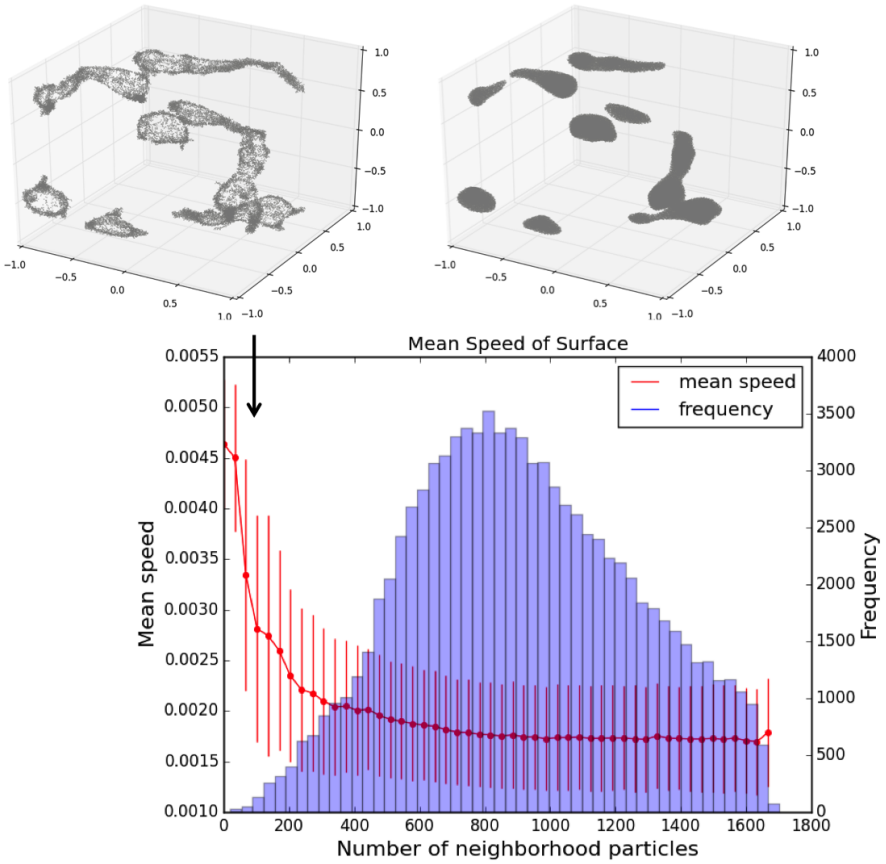

図 5.8: 大きな群れの表面と内部の速度分布

横軸は各個体の近傍粒子数を表し，縦軸はその近傍粒子数を持つ個体の平均速度を表す．この分析は，近傍粒子が少ない個体が群れの外側に，多い個体が群れの内部にいると仮定している．実際，近傍粒子数が 500 以下の個体と 500 以上の個体をプロットすると，図の上のように内と外が分離できていることがわかる．

5.4.2　既存手法による解析結果の考察

本研究の大規模なボイドシミュレーションによって，パラメータを共有しつつも全く違うダイナミクスを持つ群れが生じることがわかった．

この分析結果は，密度分布に着目するというヒューリスティクスな方法による群れの抽出によって得られたものである．しかし，今回とは違うパラメータ領域では，もっと複雑な群れが存在する可能性があり，そもそも人間が目で見て群れの存在が判別できないような場合には，このままでは対応できない．従って，なんらかの指標に基づく，機械学習を用いた群れの自動抽出ができないか検討したい．そこで検討したいのが多様体の存在である．

群れのように，全体で速度を共有したり運動の位置を共有するという現象は，各個体からみると，その自由度が拘束を受けているとみることができる．従って，群れの運動時系列の自由度は，個体の自由度の合算に比べて圧倒的に低い次元を持ち，なんらかの多様体上に分布することが期待される．実際，タンパク質のフォールディングを同様の視点でとらえ，ISOMAP をその自由エネルギー地形の可視化に用い，2 次元程度でゆらぎの 9 割以上を捉えることに成功したという報告がある [19].

本研究においても，同様に多様体学習による次元圧縮を用いて，群れの自動抽出を行うことを考えたい．そのために,「群れ」を「ある同一の多様体に，その時系列データが埋め込まれた個体の集合」と定義する．これは次のような考えに基づく．今，Nstep 分の各粒子の位置と運動量情報（6 次元）を 1 サンプルとするデータセットを考える（図 5.9）.「群れ」とは，図 5.2 で見たように，速度や位置を群れ全体で共有し全体で一つの個体のように振る舞う現象のことである．例えば ”highly parallel ”を例にとると，前述したデータセットは，群れの平均位置と平均運動量の時系列データの点を中心として 3 次元程度の多様体にデータが分布すると考えられる．なぜならば，まず速度は全ての個体で同一であるため自由度を持たない．一方，位置に関しては，群れが 3 次元空間中に大きさを持っているため，3 次元の自由度を持つ．ただし，個体の相対位置は保持されるので，時間軸方向には自由度を持たない．従って，データ分布の自由度は 3 次元となるのである．

このように考えることで,「群れ」の抽出を多様体抽出で行うことが可能になる．しかし，前に議論したように，多様体学習アルゴリズムは写像関数を生成せず，次元圧縮をしたいデータをすべて一度にメモリにのせる必要がある．このような制約は大自由度の力学系の分析には適さない．そこで本研究では深層学習を多様体学習機として用いて群れの抽出をすることを試みる．

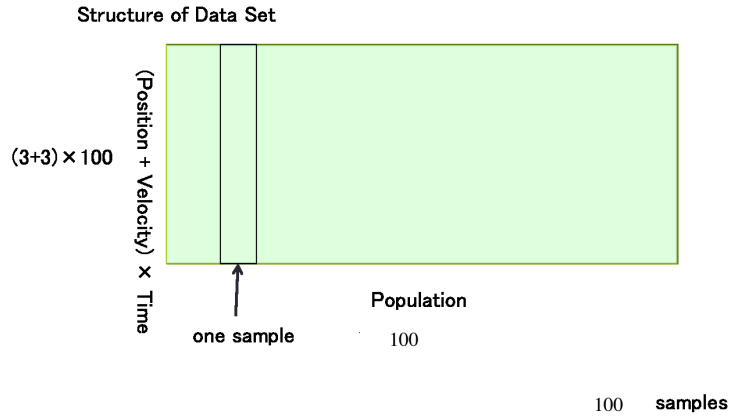

図 5.9: 多様体学習用データセット構成
100step 分の速度と位置の時系列を 1 サンプルとするデータ.

5.5 深層学習を用いた分析

5.5.1 多様体学習による群れの抽出

前節で議論したように，深層学習を用いた多様体構造の抽出によって，大自由度な系の「群れ」構造抽出を試みる．

そのためにまず，多様体学習によって群れが抽出可能であるかを検証した．具体的には，前章で紹介した 4 つのタイプ（swarm，torus，dynamic parallel，highly parallel）に対応する 100 匹程度の個体によって形成される群れを生成し，その時系列データを t-SNE によって次元圧縮した．

t-SNE への入力データセットは，100step 分の各粒子の位置と運動量情報（6 次元）を 1 サンプルとするデータセットになっており，サンプル数は 100（個体数）であった（図 5.9）．次元圧縮の結果が図 5.10 である．図にあるように，それぞれの種類の群れ毎に綺麗に分離される次元圧縮が行われることが確認された．この結果から，大自由度な系に対しても同様に，多様体学習が有用であると考える．

5.6 DNN のハイパーパラメータチューニングと特異値分布

5.6.1 目的

前章でデータに大きな摂動を加える実験でみられたように，ヤコビアンの特異値分布の急峻さと学習パフォーマンスの間に関係性があることが期待される．ま

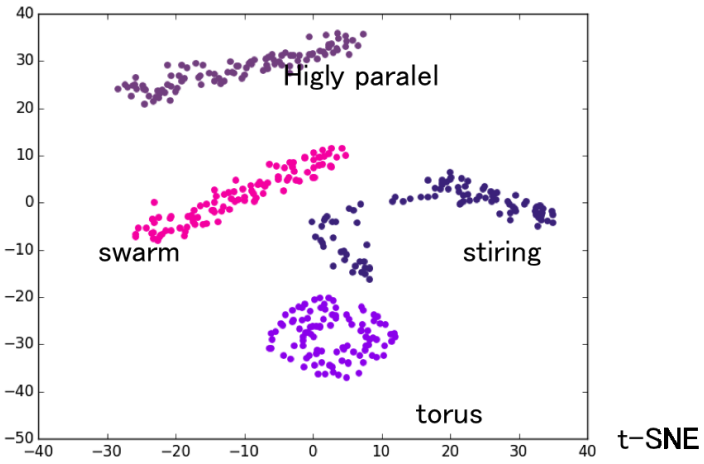

図 5.10: 多様体学習（t-SNE）による，次元圧縮の結果
各種類の群れを同時に多様体学習にかけた結果．それぞれのデータは，空間的に同じ座標にいる場合もあるが，この空間上では綺麗に分離されている．

た，教師なし学習において，多様体をうまくとらえることが高いパフォーマンスの実現に必要であると考えられることからも，多様体の次元などの情報が含まれる特異値分布を教師なし学習のハイパーパラメータチューニングの指標として利用できることが期待される．

そこで以下において，DNN の学習でよく利用される正則化項の係数というハイパーパラメータのチューニングに着目し，これと特異値分布の関係を調べた．

5.6.2 実験方法

モデルの般化能力を向上させるために，モデルを単純化するバイアス項を次のように損失関数に追加することが行われる．

$$Loss' = Loss + \lambda \sum_{i}^{M} |W_i|^q \tag{5.4}$$

W_i は，モデルパラメータを表す．DNN の場合は重みになる．特に q=1 の場合を L1 正則化といい，λ を大きくするに従っていくつかの W_i が 0 となっていき疎な解が得られる．

Deep Auto Encoder では，中間層のノード数を大きく取ると入力をそのまま出力にマッピングするような恒等写像が学習される可能性が出てくる．一方で，中間

層を減らしすぎると，DNN の表現力が大きく損なわれてしまう．多様体の次元が不明な実データの学習においては，最適な中間ノード数を設定できないため，中間層のノード数を大きく取っておいて，正則化によって恒等写像となるのを防ぐようなことがよく行われる．

しかし，その場合にも最適な λ を選ばなければならないという問題は残る．そこで本実験では，λ と特異値分布の関係を調べ λ の決定に特異値分布の情報が利用できないかの検討を行った．

学習には，これまでの実験で使用したような中間ノード数を絞ったネットワークは用いず，正則化項によるモデルの単純化を行った（図 5.11）．データセットには，前章でも使用した，10 次元空間中の 4 次元球のデータセットを用いた．従って，多様体を捉えている場合の特異値分布は，3 次元目と 4 次元目の間で急峻な減少をすることが予測される．

5.6.3 実験結果

λ の増加に応じて，auto encoder の二乗誤差（RMS）は変化した．特に λ を大きくとりすぎると DNN の表現力が落ち，RMS が増大することが確認された（図 5.12）．

同時に，DNN が恒等写像になっていないかを検証するために，トレーニングデータにはないデータセットを作成し（学習データセットとは逆向の半球のデータ），その RMS を調べた．すると，小さな λ において RMS が小さくなっていることが確認され，恒等写像的になっていることが示唆された．

以上より，恒等写像ではなく低い RMS が実現されているのは，$\lambda = 10^{-4}$ 付近であると考えられる．このときの特異値分布をみると（図 5.13），ちょうど $\lambda = 10^{-4}$ 付近を境に，これよりも λ が大きければ，多様体の次元以上に次元が圧縮され，小さければ圧縮がうまくいっていないことが観測された．そして，多様体の次元である 3 次元目と 4 次元目の特異値の減少率が最大化されるのが $\lambda = 10^{-4}$ になっていた．

これらの分析結果は，λ 等のハイパーパラメータの決定において，特異値分布をみることが有用な指標になることを示唆する．しかし，単一の実験設定での結果であり，他の条件やハイパーパラメータでどうなるかは不明である．しかしながら，これまでの実験 $1 \sim 5$ の結果を考え合わせると，多様体を捉えることがよい結果を得るための条件となっている場合，急峻な特異値分布が DNN のパフォーマンスと関係する傾向があるとはいえよう．

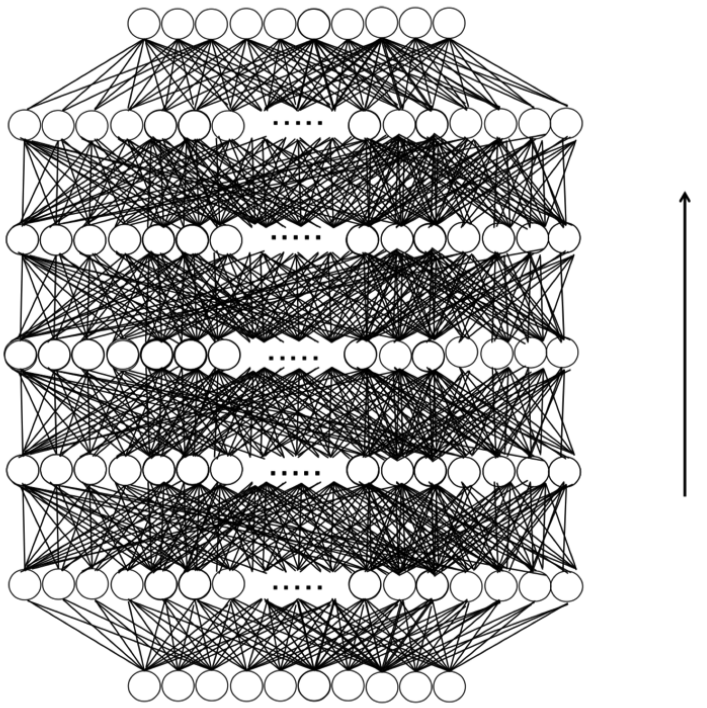

図 5.11: 使用した Deep Auto Encoder

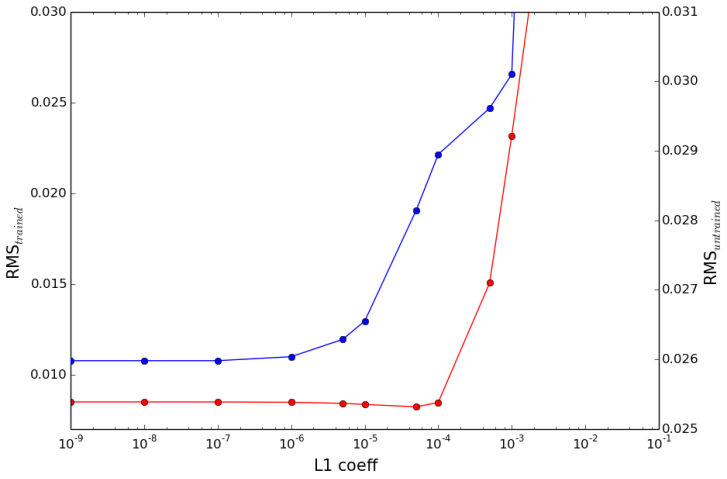

図 5.12: 2 乗誤差グラフ
赤（左軸）: 学習データセットに対する誤差．青（右軸）: 学習データとは違うデータセット（学習データとは逆方向を向いた半球）に対する誤差．この値が小さいと，恒等写像的になっている可能性がある．

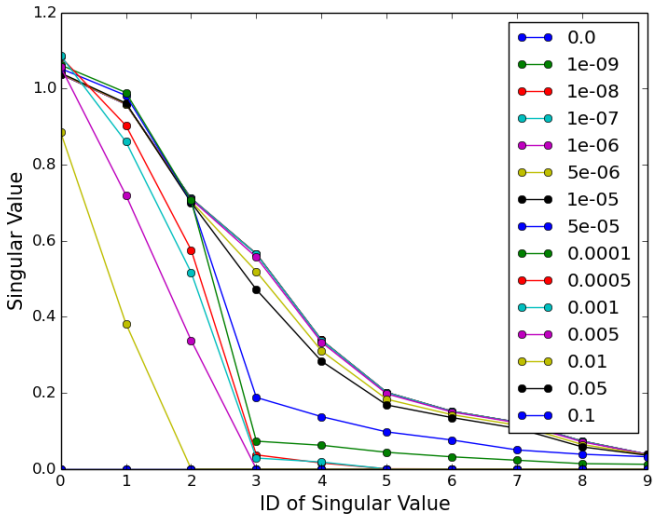

図 5.13: L1 正則化項の係数と，特異値分布（Layer3）
図 5.12 において，学習データに対する二乗誤差が小さく，かつ未知データに対して二乗誤差が大きい L1 正則化項の係数は，特異値分布が多様体の次元の前後で最も急峻な形となっている．

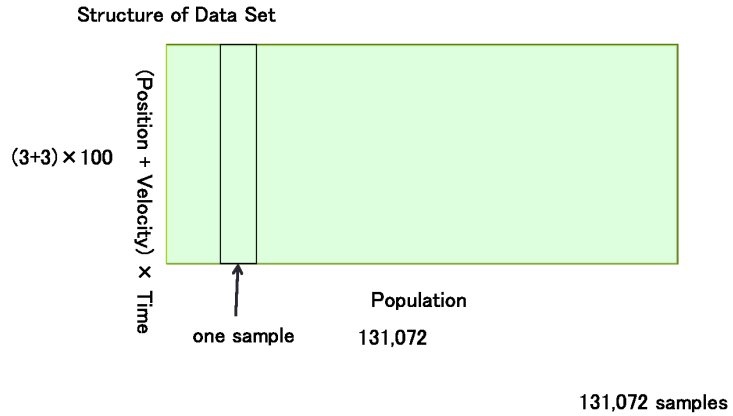

図 5.14: 深層学習用データセット構成
100step 分の速度と位置の時系列を 1 サンプルとするデータ．全部で個体数分の 131,072 個のサンプルとなる．

5.6.4 深層学習による群れ抽出

以上の結果を踏まえ DNN による群れの抽出を行う．まず，学習に用いるデータセットは，少数自由度のときと同じ各個体の 100step 分の位置と速度を 1 サンプルとする，サンプル数 131,072 のデータセット（図 5.14）である．

このデータを用いて，RBM を用いた DNN の教師なし学習を行い，その出力を K-means を用いて 15 のクラスに分類した（図 5.15）．この学習を行う際のハイパーパラメータのチューニングでは，本稿第 3 章の実験 6 の結果に基づき，より特異値分布が急峻になるようなパタメータを選ぶようにした．その結果の特異値分布が図 5.16 である．前に”highly parallel”の群れの多様体構造について考察したように,「群れ」構造は $O(1)$ 程度になることが期待される．従って，この特異値分布のグラフから多様体構造の抽出に成功していることが示唆される．

5.6.5 分類結果の分析

DNN による変換によって，距離情報だけに基づいた K-means の結果に比べて，群れがより適切に抽出できているようにみえる結果が得られた（図 5.17）．特に中央の大きな群れが K-means 法のみの場合では分離されてしまっている点に注目されたい．K-means の初期値によって，これはある程度変動するが，同じ群れが分離されてしまう可能性があるということに変わりはない．

最後に，得られた「群れ」が DBSCAN によって抽出された「群れ」に対応するか検証するため，群れのエネルギー分布をみた．その結果，小さな群れはエネル

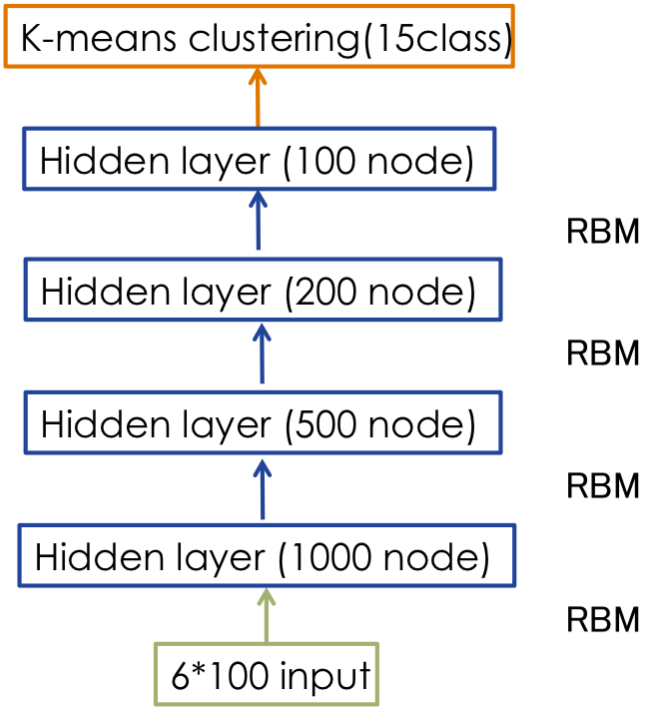

図 5.15: 学習に用いた Deep Belief Networks

バイナリ RBM を用いている．また，次元が圧縮されるよう，層をへる毎にノード数を減らすように工夫している．そして，この DNN の出力結果を集め，K-means にかけて出てきたクラスタを群れとした．

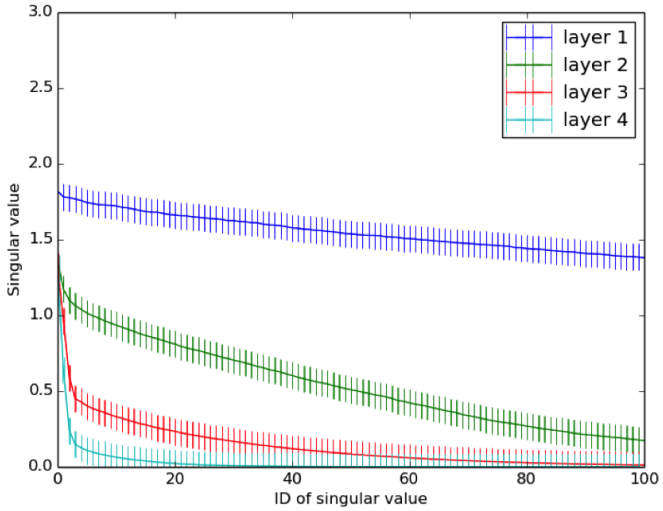

図 5.16: 学習後の特異値分布
最終層において急峻な特異値分布が実現されている．エラーバーは標準偏差をあらわす．

ギーが大きい一方，大きな群れではエネルギーが小さくなるという，前章と同様の結果（図 5.7）が得られ（図 5.18），予想された群れの抽出ができていることが確認された．

以上から，深層学習によって群れの自動抽出ができたことが確認された．

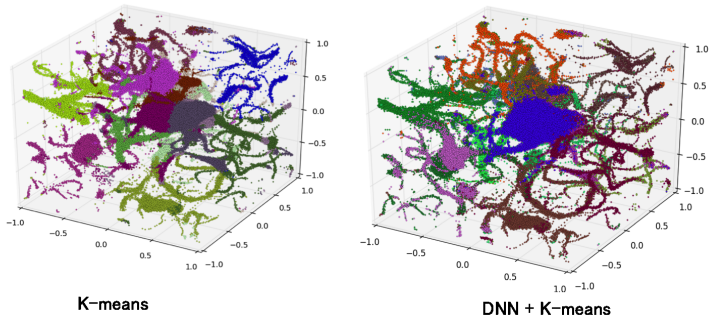

図 5.17: 左：K-means によるクラスタリング結果　右：深層学習を併用した結果
K-means の結果と比較して，DNN の結果は，大きな群れを綺麗に抽出できている点や，フィラメントと大きな群れが混じったクラスタがない点などが優っている．

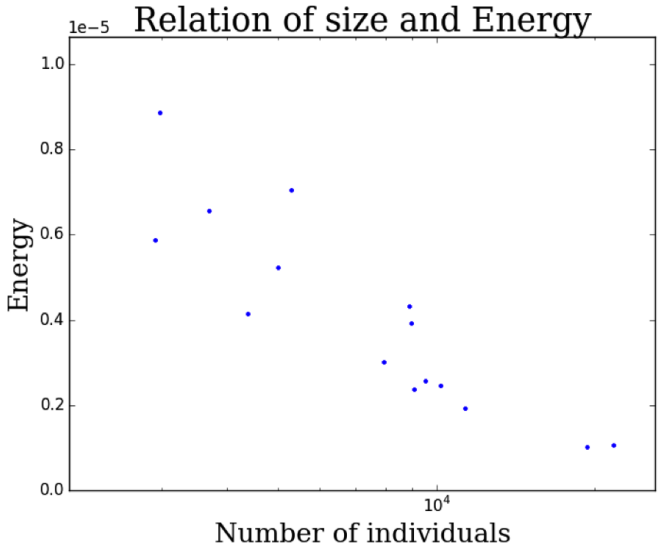

図 5.18: 深層学習によって抽出された群れのエネルギー分布
前章の既存手法による結果（図 5.7）と同様に，大きな群れのエネルギーが小さいことが確認される．

5.7　まとめと考察

本章では，画像データ以外に多様体構造を持ったデータセットとして，大自由度ボイドモデルに着目しその分析に DNN を適用した．

具体的には，未だ多くの研究が行われていない複数の「群れ」の間の相互作用ダイナミクスを解明することを目的に，大規模な集団運動のシミレーションにおいて生成される複雑なパターンから「群れ」構造を抽出することを試みた．本研究では，ヒューリスティクスによってもなんとか「群れ」の抽出が行えたものの，このような系では性質の違う複数の「群れ」が複雑なパターンを形成すると考えられるため，これまでのようにヒューリスティクスによる「群れ」の抽出では限界が生じてくると考えられる．そこで本研究では,「群れ」を「ある同一の多様体に，その時系列データが埋め込まれた個体の集合」と定義し，DNN でこの構造を抽出することを行った．DNN を用いたのは，多様体学習アルゴリズム等の既存手法ではメモリ容量の制約から，大規模なボイドシミュレーションの分析が難しいためである．分析の結果，ヒューリスティクスによって事前に抽出していた群れと同様の性質を持った「群れ」の抽出に成功した．この「群れ」の定義は，今回使用した以外のモデルや実際の群れの観測データにも適用できる一般的な定義であると考えられ，より複雑で大規模な集団運動にアプローチしようとする場合に有用なものであると考えらえれる．このように「群れ」を定義し，その抽出に成功した研究はこれまでにない．

また，これと関連して，DNN の学習時に特異値分布の形状をハイパーパラメータ決定の指標として用いることを検討し，特異値分布が最も急峻になるようにハイパーパラメータを選ぶことで高いパフォーマンスが得られることが，人工データでの検証と実際のボイドモデルの分析において確認された．特に，予想されたボイドモデル時系列データセットの多様体の次元と，特異値分布から予想される次元がおおよそ一致したことは，特異値分布の情報を，ハイパーパラメータチューニングの指標としてだでなく，対象となる力学系の性質を知る為の有用な手段とできることを示唆する．このように，特異値分布をハイパーパラメータチューニングの指標とできることは，一般に困難な [8]DNN のハイパーパラメータチューニングにおいて，具体的に DNN のどの層がパフォーマンスを下げる要因になっているか等の詳細な検討を可能にするため，非常に有用であると考えられる．また，Constractive Auto Encoder のように，DNN の学習自体にヤコビアンの情報を利用するものはあるものの，ハイパーパラメータチューニングの指標としてこれを利用した研究はこれまでにない．

一方，既存手法と DNN のどちらでも抽出に成功したこの「群れ」を元に，それぞれの群れ毎の分析を行ったところ，速度が遅く淀んでおりその外側と内側で速

度が違うような複雑な構造を持った群れと，その周りを飛び交う速度の速いフィラメント状の群れが観測された．これは，複数の群れが相互作用するような大規模なボイドシミュレーションならではの現象であると考えられ，一つの群れに着目した多くの既存研究では扱ってこなかった現象である．少ないものの，このような複数の群れが相互作用するようなモデルはこれまでにも研究されているが [19]，それらの研究では個体の内部パラメータの違いで違う種類の群れを生成しており，本研究のように同じパラメータにも関わらずこのような複雑な群れの構造が生成されたことは，これからの集団運動の研究に対して新しい知見を与えれらたものと考える．

第6章　議論と結論

6.1　総合考察

本研究の目的は，データセットの構造をありのままに観察することであり，その第一歩として，DNN を用いることで具体的にいくつかのデータセットの幾何構造の観測を試みた．

そのためにまず，DNN がデータセットの観測装置として有効であるかの検討を行った．具体的には，多様体仮説

- データセットは，そのデータの次元 d より十分に低い次元 d_M を持つ（微分可能）多様体上に埋め込まれれている．

を満たすと考えられるデータセットを用いて，

- パフォーマンスの高い学習済み Deep Neural Networks（以下，DNN）は，上の多様体を多様体と同じ次元の大域的な座標系へ写像する機能をもつ．

という，DNN を観測装置として利用する上で必要な仮説の検証を行った．

具体的には，多様体と同じ次元の中間層を持つ Deep Auto Encoder に 10 次元空間中の n 次元球を学習させその写像関数を分析した．その結果，その写像関数のヤコビアンの特異値分布の分析によって DNN が多様体の次元を捉えていることを，右特異ベクトルの分析によって DNN が接空間をおおよそ捉えていることを確認した．一方で，本研究でデータセットの構造として設定した半球のような，正方形で引き伸ばしなしに被覆できない構造がある場合は，右特異ベクトルの方向と多様体の接線方向の間に誤差が生じることも確認された．データセットの観測に DNN を用いる場合は，この誤差の存在を常に意識する必要がある．

これまでの研究では，可視化に頼った方法によってこの仮説を確認しており [30][71]，定量的な手法による検証は本研究が初めてである．またこれと関連して，可視化によって検証が行えない高次元空間中の n($n \geq 3$) 次元多様体での検証を行ったことも本研究の貢献となる．また，前述したような半球の切断面付近での誤差は入江らの研究 [71] でも確認されていたものの，入江らはこれを無視していた．本研究ではその誤差について検討し，前述したようにこの誤差の原因として

幾何学的な要因を提示することができた．この現象が生じる根底には，DNN の写像関数が同相的な変換をしていることがあると考えられる．従って，本研究で示された誤差要因から，正方形では伸縮なしに被覆できないような多様体構造を持つデータセットを学習する為に，あえて DNN 写像関数の同相性を壊すような機構を導入することで，よりパフォーマンスの高いアルゴリズムが開発できる可能性が示唆されたと考える．今後，この点を検討していきたい．

次に本研究では，実践的なデータセットである MNIST データセットによってトレーニングされた DNN も仮説を満たす写像機能を持つかについて，定量的な検証を行った．MNIST データセットは，hinton らの研究 [30] 等によって多様体構造の存在と，その次元が $O(1)$ 程度になることが示唆されている．一方で，これらの研究は中間層が 2 となる Deep Auto Encoder を用いることで可視化によって検証を行っており，十分な検証がなされているとは言い難い．また，MNIST の識別においてパフォーマンスが高い DNN では，このような極端に小さな中間層をもつネットワーク構造は用いられていない．そこで本研究では実際に使用されるような中間層がより高次元な DNN を定量的に分析し，実践的な状況においてもパフォーマンスの高い DNN が仮説を満たすことを確認することを試みた．

その結果，pre-training（教師なし学習）後の特異値分布の分析から DNN が捉えている多様体の次元が $O(1)$ になることが確認され，右特異ベクトルの分析から，DNN が多様体の接空間を捉えていることが，左特異ベクトルの分析から DNN がそれを大域的な座標系へ展開していることが確認された．

さらに，fine-tuning（教師あり学習）後の特異値分布の分析によって，fine-tuning（教師あり学習）が教師なし学習とは違うダイナミクスをもつこともわかった．特に fine-tuning 後の特異値分布が全体的に増大することから，fine-tuning によって DNN に情報が追加されていると考えられる．これらの結果は，pre-training 時のような教師なし学習の際には多様体構造を大域的な座標系へ変換するような単射的（n 対 n 写像）な写像機能が獲得され，fine-tuning 時のような教師あり学習では，違うクラスの部分多様体の間に存在するような判別の難しいデータが所属するクラスについての情報を与えるように，クラス間に分離面を引くような写像機能（n 対 m 写像：m<<n）が獲得されると解釈可能であると考える．一方で，fine-tuning 後の特異値分布や右特異ベクトルが，pre-training 後と比較して全く違うものとはなっていないこと，及び，pre-training が教師あり学習に対しても有用であることから，教師あり学習時においても多様体構造の情報が有用に利用できると考えられる．つまり，DNN は，これら 2 種類のダイナミクスの相互作用によって入力空間の写像を行っていると考えられる．

また，クラス毎の特異値分布の形状の違いを PCA を用いた次元圧縮を通して検討したところ，特異値分布がクラス毎にそれぞれ違う形状をもつことが確認され

た．このことは，データセットの意味的なクラスと幾何的な構造が関係することを示唆しており，MNIST データセットが多様体仮説を満たすことを強く支持する．幾何構造と意味的なクラスの間の関係性をこのように直接的に示した研究はこれまでにない．

これらの結果をふまえ ImageNet データセットの観察を行った．その結果，高次の層において少数の大きな特異値と大多数のほぼ 0 の特異値という，急峻な特異値分布が得られた．また，特異ベクトルをみた結果，特異値の大きいベクトルでは各入力データに固有のようにみえる feature を捉えるベクトルがみられた一方で，特異値の小さなベクトルでは，空間的に広く分布したノイズのような構造がみられた．これは MNIST でみられたものと同様の構造である．これらの結果は，この DNN が多様体の接空間の間の写像関数を獲得していることと，ImageNet データセットが多様体構造をもつことを示唆する．

さらに，DNN が多様体の接空間の写像を行っていることをより直接的に確認することを試みた．具体的には，次のようなことを行った．もし，1 より大きい特異値に対応する特異ベクトルが多様体の接空間を表しているとすれば，入力データにその方向の摂動を加えたとしてもニューラルネットワークの出力はあまり変動しないはずである（クラスが変わらないため）．一方，多様体の垂直方向である 1 より小さい特異値に対応する特異ベクトル方向への摂動を加えると出力は大きく変動するはずである．これを確かめたところ，実際にニューラルネットワークの出力が，接線方向の摂動に対して相対的にロバストであることが確認された．このことは，ImageNet データセットが多様体構造をもっていること，及び AlexNet がその接空間を入力空間から抽出するような写像関数を学習していることを意味する．

次に，多様体から離れるような摂動と特異値分布の関係を調べた．そのために入力画像にノイズを付加した上で特異値の算出を行なった．その結果，高次の層においてノイズの増大に応じて特異値分布全体が小さくなっていくことが観察された．また，このときの出力は不正解となっていた．これは，データ点が多様体から大きく離れた場合，DNN 内でうまく情報が伝達されなくなることを意味していると考えられる．この結果は，学習がうまくいかない状態を特異値分布で捉えることができる可能性を示唆すると考えられ，DNN の大自由度力学系への応用では，後述するようにこの特異値分布とパフォーマンスの関係を用いたハイパーパラメータチューニングに挑戦した．

最後に，MNIST データセットを用いた分析と同様にクラス毎の特異値分布をみた．MNIST と同様に特異値をベクトルとみなして主成分分析を行うと，層を経るにつれてクラス毎にクラスタを作る傾向があるようにみえた．特に，意味的な階層が低次なクラスは，より低次の層でクラスタを形成しているようにみえた．

これらの主観的な観察を踏まえ，この結果を定量的に検証した．具体的には，良いクラスタの指標として使用される，クラスタの大きさを表す標準偏差（より良いクラスタは狭い領域に分布する）と，クラスタ間の平均距離（より良いクラスタはお互いに離れている）を用いて上の観察結果を検証した．この分析の結果，低次のクラスは layer1 からランダムにクラスタを生成した場合と比較して有意に小さな標準偏差を持つことがわかった．一方，高次のクラスにおいては layer5 のみでランダムな場合との間に有意な差をもつとわかった．一方，クラスタ間平均距離は，低次のクラスでも高次のクラスでも全ての層でランダムな場合と比較して，有意に離れたクラスタが形成されることが確認された．また高次のクラスタにおいて holm 法を用いた多重比較を行ったところ，クラスタ間平均距離に関して layer5 と layer2,3 の間に優位な差がみられた．このことは，layer5 でのクラスタ間平均距離がその他の層に比べて大きい傾向があることを示唆する．

これらの結果から，低次のクラスでは layer1 からクラス毎にクラスタが形成される傾向がある一方で，高次のクラスでは layer5 になって初めてクラスタが形成される傾向があるということが言える．これは，データセット分布の幾何構造とその意味的階層構造が関係することを示唆する．ただし，これらの結果を確定するには，画像数を増やしより多くのサンプル数で検証することや，より深い層等の違うネットワーク構造や，違うハイパーパラメータで学習された DNN での検証が必要であると考えられる．しかしながら，幾何的な構造と意味的な階層構造の関係を示唆する結果を得たことは大きな第一歩であると考えられる．

これらが正しい場合，特に文化的要因や学習によらない人間の低次の表象の形成において，データの何らかの幾何学構造が重要な意味を持つことになり，その場合，本研究の結果が認知科学等の領域に与える影響は大きいものと考えられる．

DNN とデータの意味的な構造の対応は，これまでにも検討されてきた．例えば，Donahue ら [22] や　Aubry ら [3] は，t-SNE や PCA を用いることで AlexNet の中間層の次元を圧縮し，そこで実現されているおおよそのデータセットの分布を取得し，クラスタの形成を議論している．また，AlexNet と別の DNN を結合させること等で，本来は生成モデルをもたない AlexNet の生成モデルを形成 [52][67] し，各層や各ノード等が担当する情報の可視化を行うことも試みられており，DNN と階層的な表象の関係が活発に議論されている．しかしながら，多様体仮説との関係性についての直接的・定量的な検討はこれまでに行われてこなかった．本研究で得られた結果は多様体仮説と意味的な階層構造の関係を示唆しており，これはこれまでになかった知見である．

次に本研究では，このような分析結果を説明するものとして,「データセットはデータの意味の階層に対応した幾何学的な階層を持つ」という仮説を提案し，その妥当性を考察した．具体的には階層的な幾何構造を持ったデータセットでトレーニ

ングされたDNNを分析し，DNNが階層的な幾何構造を持ったデータセットを大域的な座標系へ写像するような学習が可能なことを示した．さらに，Decoder層の可視化より，DNNが多様体を折り紙を展開するように段階的に変換しながら，階層的な構造を持った多様体を大域的な座標系へ写像しているとい描像を得た．しかしながら，Encoder層の写像関数の分析による検証では，これを強く支持する結果は得られなかった．この仮説の妥当性と，DNNが階層的な幾何構造を段階的に展開するという知見の正しさについては，さらなる検証が必要であると考える．

ところで，本研究のようにDNNを用いない場合でも，3層パーセプトロンの中間層を大きくすれば有界閉集合上の任意の連続関数を任意の制度で近似できることがわかっている [26][17]．実際に，DNNが学習した写像関数を3層パーセプトロンで近似できることもわかっている [5]．このような現状があるため，なぜ多層化することが有用であるのかという基本的な問いかけは，未だに明確な解答を得ているとは言い難い．本研究で提示した上の描像は，この問題に対して示唆を与える．なぜならば，以下のように入れ子状の関数によって複雑な表現力を得ていくDNNは，折り紙のようなデータ構造をはじめからネットワークとしてモデル化しているといえるからである．

$$z(x) = f_n(f_{n-1}(\cdots f_0(x)\cdots)) \tag{6.1}$$

つまり，ネットワーク構造としてのデータのモデル化がDNNが有効に機能する理由の一つになるのではないかと考えられる．

ところで，多くの先行研究において，DNNによって抽象的なカテゴリ情報が形成されることが示唆されている [40] [39]．このような抽象化は，本研究で示した多様体仮説とDNNによるその写像機構だけでは説明できない．なぜならば，多様体の大域的な座標系への写像は単射であるからである．抽象化のためには，複数の入力データを1つのカテゴリに対応させるようなn対1写像も必要になると考えられる．一方で，本研究の第3章で示したように，教師あり学習によるfine-tuning後のDNNの内部状態はpre-trainig後と違う様相を示していた．このことから考えると，DNNの内部では，多様体を捉えてそれを段階的に展開していく単射性を持ったダイナミクスと，複数のデータを1つのラベル（ノード）に対応付ける教師あり学習や，領域のパターンを1つの値に代表させるpooling（付録C）のようなn対1写像を行うダイナミクスの2種類の機構が働いており，これらの結果として情報の抽象化が行われるものと考えられる．ただし，ここで多様体を捉える写像が情報の抽象化に貢献しないとは考えない．なぜならば，階層構造に応じて多様体が展開されることで初めて，n対1写像すべき領域が一つの連結した領域として出現すると考えられるためである．そしてこの連結した領域が本研究で示した仮説のように，カテゴリの階層性に応じて段階的に展開されるならば，それに応じ

たn対1写像によって，情報の意味的な階層に応じた情報の抽象化が行えることを意味し，その結果として概念に対応する抽象構造が実現されるものと考えられる．

本研究では，Artificial Neural Networks（以下，　ANN）を用いて，それが多様体構造を大域的座標系へ写像する機能を有するかを検討してきた．この機能は，ANNだけでなく人間も有する可能性もあると考えられる．なぜならば，データセットが多様体構造を持つならば，この構造を抽出できる生物種はデータの内容毎に特別な前提をおく必要がなく，より容易に高次元データから情報の抽出を行うことが可能になるからである．実際に，デカルロらは，人間の視覚系が同様の処理を行っているという仮説を提唱している [21].

次に，画像データセット以外で多様体仮説が成立している対象として大自由度力学系の時系列データに着目し，その分析にDNNを適用した．具体的には，未だ多くの研究が行われていない鳥や動物などの複数の「群れ」の間の相互作用ダイナミクスを解明することを目的に，大規模な集団運動のシミレーションにおいて生成される複雑なパターンから「群れ」構造を抽出することを試みた．このような系では性質の違う複数の「群れ」が複雑なパターンを形成するため，これまでのようにヒューリスティクスによる「群れ」の抽出では限界が生じてくると考えられる．そこで,「群れ」を「ある同一の多様体に，その時系列データが埋め込まれた個体の集合」と定義し，DNNでこの多様体構造を抽出することで「群れ」の抽出を試みた．DNNを用いたのは，多様体学習アルゴリズム等の既存手法ではメモリ容量の制約から，大規模なボイドシミュレーションの分析が難しいためである．分析の結果，ヒューリスティクスによって事前に抽出していた群れと同様の性質を持った「群れ」の抽出に成功した．この「群れ」の定義は，今回使用した以外のモデルや実際の群れの観測データにも適用できる一般的な定義であると考えられ，より複雑で大規模な集団運動にアプローチしようとする場合に有用なものであると考えらえれる．このように「群れ」を定義し，その抽出に成功した研究はこれまでにない．

また，これと関連して，DNNの学習時に特異値分布の形状をハイパーパラメータ決定の指標として用いることを検討し，特異値分布が最も急峻になるようにハイパーパラメータを選ぶことで高いパフォーマンスが得られることが，人工データでの検証と実際のボイドモデルの分析において確認された．特に，予想されたボイド時系列データの多様体次元と特異値分布から予想される次元がおおよそ一致したことは，特異値分布の情報を，ハイパーパラメータチューニングの指標としてだでなく，対象となる力学系の性質を知るための手法とできることを示唆する．このように，特異値分布をハイパーパラメータチューニングの指標とできることは，一般に困難な [8]DNNのハイパーパラメータチューニングにおいて，具体的にDNNのどの層がパフォーマンスを下げる要因になっているか等の詳細な検討を

可能にするため，非常に有用であると考えられる．Constractive Auto Encoder のように，DNN の学習自体にヤコビアンの情報を利用するものはあるものの，ハイパーパラメータチューニングの指標としてこれを利用した研究はこれまでにない．

一方，既存手法と DNN のどちらでも抽出に成功したこの「群れ」を元に，それぞれの群れ毎の分析を行ったところ，速度が遅く淀んでおりその外側と内側で速度が違うような複雑な構造を持った群れと，その周りを飛び交う速度の速いフィラメント状の群れが観測された．これは，複数の群れが相互作用するような大規模なボイドモデルならではの現象であると考えられ，一つの群れに着目した多くの既存研究では扱ってこなかった現象である．また，少ないもののこのような複数の群れが相互作用するようなモデルはこれまでにも研究されているが [19]，その研究では個体の内部パラメータの違いで違う種類の群れを生成しており，本研究のように同じパラメータにも関わらずこのような複雑な群れの構造が生成されたことは，集団運動の研究に対して新たな知見を与えれらたものと考える．

6.2 今後の展望

本研究の分析は，その計算量の多さのため，比較的少ないサンプル数で行なわれた．従って，より確かな結果を得るためには，より高度な計算機環境を利用したサンプル数の増大が必要だと考えられる．

また近年，本研究で分析したような DNN モデル以外に，多くのユニークなモデルが提案されている．例えば，DCGAN（Deep Convolutional Generative Adversarial Networks）と呼ばれる DNN は，本物の画像で見分けがつかないような画像生成モデルを実現している [45]．また，DNN が音声と映像といった違うモダリティを統合する学習に有用であることも示されており，このような現象と多様体仮説との関係についても検討したい．このように新しいモデルを解析することで，データセットのもつ幾何学的構造についてのさらなる知見の収集を行う予定である．

また，位相的機械学習と呼ばれる，多様体の穴の構造に注目してパーシステントホモロジーと呼ばれる量等を取得する手法があり [72]，これをデータセットの構造の理解に応用していきたい．ただし，この手法には次元の増大に応じて大きな計算コストがかかり，今回分析した Image Net のような 10 万オーダーの次元のデータセットには適用できない．そこで，DNN の写像の同相性等を検討しながら，DNN で次元圧縮したデータセットの構造に対して位相的機械学習法を適用するといったようなことを行っていきたい．

Publications

Refereed papers

- Yhoichi Mototake and Takashi Ikegami, The dynamics of deep neural networks, Proceedings of the Twentieth International Symposium on Artificial Life and Robotics, 20. January 2015.

- Y.Mototake, and T.Ikegami A Simulation Study of Large Scale Swarms, SWARM 2015,pp.446-450, Kyoto university, Oct. 28-30, 2015.

Unrefereed papers

- 本武 陽一, 池上 高志 : Deep Neural Networks の内部構造の解析とその応用（ポスター）, The 18th Information-Based Induction Science Workshop (IBIS'15), November 2015.

- 本武陽一，池上高志 : Deep Neural Networks の力学的解析, 2015 年度人工知能学会全国大会（第 29 回）, 北海道, 2C3-OS-06b-4in,2015.

- 本武陽一，池上高志 : 大自由度ボイドモデルの解析, 日本物理学会第 70 回年次大会, 22aAJ-4,2015.

- 本武 陽一, 岡 瑞起, 池上 高志 : ディープニューラルネットワーク内ダイナミクスの力学的解析, 2014 年度人工知能学会全国大会（第 28 回）, 愛媛, 3H4-OS-24b-4, 2014.

References

[1] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines*. *Cognitive science*, Vol. 9, No. 1, pp. 147–169, 1985.

[2] Ichiro Aoki. A simulation study on the schoolng mechanism in fish. *Bull. Japan. Soc. Sci. Fish.*, Vol. 48, , 1982.

[3] Mathieu Aubry and Bryan C Russell. Understanding deep features with computer-generated imagery. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2875–2883, 2015.

[4] Bjørn Erik Axelsen, Leif Nøttestad, Anders Fernö, Arne Johannessen, and Ole Arve Misund. ‘ await ’in the pelagic: dynamic trade-off between reproduction and survival within a herring school splitting vertically during spawning. 2000.

[5] Jimmy Ba and Rich Caruana. Do deep nets really need to be deep? In *Advances in Neural Information Processing Systems*, pp. 2654–2662, 2014.

[6] HB Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? 1972.

[7] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, Vol. 15, No. 6, pp. 1373–1396, 2003.

[8] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*, pp. 437–478. Springer, 2012.

[9] Yoshua Bengio, Aaron Courville, and Pierre Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 35, No. 8, pp. 1798–1828, 2013.

[10] Eric Bertin, Michel Droz, and Guillaume Grégoire. Hydrodynamic equations for self-propelled particles: microscopic derivation and stability analysis. *Journal of Physics A: Mathematical and Theoretical*, Vol. 42, No. 44, p. 445001, 2009.

[11] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, Vol. 107, No. 26, pp. 11865–11870, 2010.

[12] Lawrence Cayton. Algorithms for manifold learning. *Univ. of California at San Diego Tech. Rep*, pp. 1–17, 2005.

[13] O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-supervised Learning*. MIT Press, 2006.

[14] Hugues Chaté, Francesco Ginelli, Guillaume Grégoire, Fernando Peruani, and Franck Raynaud. Modeling collective motion: variations on the vicsek model. *The European Physical Journal B*, Vol. 64, No. 3-4, pp. 451–456, 2008.

[15] Garrison W Cottrell. New life for neural networks. *networks*, Vol. 5, p. 6, 2006.

[16] Iain D Couzin and Jens Krause. Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, Vol. 32, pp. 1–75, 2003.

[17] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, Vol. 2, No. 4, pp. 303–314, 1989.

[18] Andras Czirok, Evan A Zamir, Andras Szabo, and Charles D Little. Multicellular sprouting during vasculogenesis. *Current topics in developmental biology*, Vol. 81, pp. 269–289, 2008.

[19] Payel Das, Mark Moll, Hernan Stamati, Lydia E Kavraki, and Cecilia Clementi. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proceedings of the National Academy of Sciences*, Vol. 103, No. 26, pp. 9885–9890, 2006.

[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and*

Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255. IEEE, 2009.

[21] James J DiCarlo and David D Cox. Untangling invariant object recognition. *Trends in cognitive sciences*, Vol. 11, No. 8, pp. 333–341, 2007.

[22] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013.

[23] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, Vol. 96, pp. 226–231, 1996.

[24] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[25] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, Vol. 36, No. 4, pp. 193–202, 1980.

[26] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural networks*, Vol. 2, No. 3, pp. 183–192, 1989.

[27] Charles G Gross. Genealogy of the " grandmother cell ". *The Neuroscientist*, Vol. 8, No. 5, pp. 512–518, 2002.

[28] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, Vol. 53, No. 2, pp. 217–288, 2011.

[29] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, Vol. 18, No. 7, pp. 1527–1554, 2006.

[30] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507, 2006.

[31] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, Vol. 148, No. 3, pp. 574–591, 1959.

[32] Andreas Huth and Christian Wissel. The simulation of the movement of fish schools. *Journal of theoretical biology*, Vol. 156, No. 3, pp. 365–385, 1992.

[33] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. In *Journal of Machine Learning Research*, pp. 695–709, 2005.

[34] Arthur Koestler, 信弥小尾, 博木村. ヨハネス・ケプラー：近代宇宙観の夜明け. The watershed. 河出書房新社, 東京, Japan, 1971.8 1971.

[35] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[36] Quoc V Le. Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 8595–8598. IEEE, 2013.

[37] Y Lecun and C Cortes. The mnist database of handwritten digits, 2009, 2009.

[38] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[39] Honglak Lee, Chaitanya Ekanadham, and Andrew Y Ng. Sparse deep belief net model for visual area v2. In *Advances in neural information processing systems*, pp. 873–880, 2008.

[40] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 609–616. ACM, 2009.

[41] Mikio Nakahara. *Geometry, topology and physics*. CRC Press, 2003.

[42] Vijay Narayan, Narayanan Menon, and Sriram Ramaswamy. Nonequilibrium steady states in a vibrated-rod monolayer: tetratic, nematic, and smectic correlations. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2006, No. 01, p. P01005, 2006.

[43] Hariharan Narayanan and Sanjoy Mitter. Sample complexity of testing the manifold hypothesis. In *Advances in Neural Information Processing Systems*, pp. 1786–1794, 2010.

[44] Akira Okubo. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, Vol. 22, pp. 1–94, 1986.

[45] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[46] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, Vol. 21, No. 4, pp. 25–34, 1987.

[47] Salah Rifai, Yann N Dauphin, Pascal Vincent, Yoshua Bengio, and Xavier Muller. The manifold tangent classifier. In *Advances in Neural Information Processing Systems*, pp. 2294–2302, 2011.

[48] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 833–840, 2011.

[49] Pernille Rørth. Collective guidance of collective cell migration. *Trends in cell biology*, Vol. 17, No. 12, pp. 575–579, 2007.

[50] Lawrence K Saul and Sam T Roweis. An introduction to locally linear embedding. *unpublished. Available at: http://www. cs. toronto. edu/˜ roweis/lle/publications. html*, 2000.

[51] Patrice Simard, Bernard Victorri, Yann LeCun, and John Denker. Tangent prop-a formalism for specifying selected invariances in an adaptive network. In *Advances in neural information processing systems*, pp. 895–903, 1992.

[52] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[53] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014.

[54] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.

[55] Theano Development Team. Deep Belief Networks Tutorial. `http://deeplearning.net/tutorial/DBN.html`. [Online; accessed 15-February-2016].

[56] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, Vol. 290, No. 5500, pp. 2319–2323, 2000.

[57] Vikrant Singh Tomar and Richard C Rose. A family of discriminative manifold learning algorithms and their application to speech recognition. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, Vol. 22, No. 1, pp. 161–171, 2014.

[58] Vikrant Singh Tomar and Richard C Rose. Manifold regularized deep neural networks. In *INTERSPEECH*, pp. 348–352, 2014.

[59] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, Vol. 9, No. 2579-2605, p. 85, 2008.

[60] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, Vol. 75, No. 6, p. 1226, 1995.

[61] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics Reports*, Vol. 517, No. 3, pp. 71–140, 2012.

[62] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, Vol. 23, No. 7, pp. 1661–1674, 2011.

[63] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103. ACM, 2008.

[64] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research*, Vol. 11, pp. 3371–3408, 2010.

[65] Kilian Q Weinberger, Benjamin D Packer, and Lawrence K Saul. Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization. In *Proceedings of the tenth international workshop on artificial intelligence and statistics*, pp. 381–388. Citeseer, 2005.

[66] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.

[67] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Computer vision–ECCV 2014*, pp. 818–833. Springer, 2014.

[68] Chiyuan Zhang, Georgios Evangelopoulos, Stephen Voinea, Lorenzo Rosasco, and Tomaso Poggio. A deep representation for invariance and music classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 6984–6988. IEEE, 2014.

[69] 竹内一将, 及川典子, 稲垣紫緒, 坂上貴洋, 和田浩史. Self-organizaition and dynamics of active matter: 研究会の私的会議録, 雑感, その他. 2009.

[70] 中山英樹. 招待講演 深層畳み込みニューラルネットによる画像特徴抽出と転移学習 (音声). 電子情報通信学会技術研究報告= IEICE technical report: 信学技報, Vol. 115, No. 146, pp. 55–59, 2015.

[71] 入江文平, 川人光男. 多層パーセプトロンによる内部表現の獲得. 電子情報通信学会論文誌 D, Vol. 73, No. 8, pp. 1173–1178, 1990.

[72] 裕章, 平岡. タンパク質構造とトポロジー: パーシステントホモロジー群入門. 共立出版, 2013.

もし私に 金と銀の光が 縫い込まれた 天の布があったなら
夜と薄明と昼を表す 漆黒と灰色と空色をした 天の布があったなら
その布を あなたの足下に広げましょう
しかし、貧しい私には、夢しか無かったのです
だから、あなたの足下に夢を
そっと歩いて欲しい、私の大切な夢だから

イエーツ

付 録A　Deep Auto Encoder

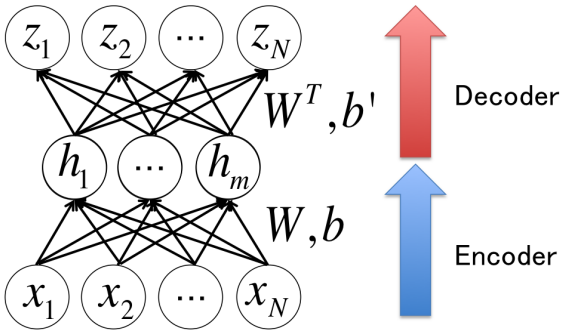

図 A.1: Auto Encoder

Deep Auto Encoder は，Auto Encoder を多層に拡張したニューラルネットワークである．そこでまず，Auto Encoder について解説する．

Auto Encoder は，図 A.1 のように入力ノード数と出力ノード数が一致し，中間層のノード数が小さな砂時計型のネットワーク構造を持つ 3 層パーセプトロンである．このネットワークの学習目的は，学習データセットによって与えられる入力 x_i と，ネットワークの出力 Z_j を一致させることである．その際の損失関数とし

ては，例えば以下のような二乗誤差関数，

$$E(\mathbf{W}) = \sum_{n=1}^{N} |x_n - z_n(\mathbf{x})| \tag{A.1}$$

$$z_k = f(\sum_{j=0}^{m} W_{jk}^T h_j + b'_k) \tag{A.2}$$

$$h_j = f(\sum_{i=0}^{N} W_{ij} x_i + b_j) \tag{A.3}$$

が用いられる．

Deep Auto Encoder は，Auto Encoder と同様に入出力関係の一致を目標として学習を行う．効率的な学習を行う為に，Deep Auto Encoder の学習には pre-training と呼ばれる layer greedy wise な方法が利用される．具体的には，図 A.2 のように，Deep Auto Encoder の各層を単独の Auto Encoder として学習しておき，その結果を組み合わせて事前学習された Deep Auto Encoder を構築する．その後，全体の最適化（fine-tuning）を行う．ちなみに，pre-training 時の第 $n+1$ 層目の Auto Encoder の学習には，学習済みの第 n 層の出力パターンが入力データとして用いられる．

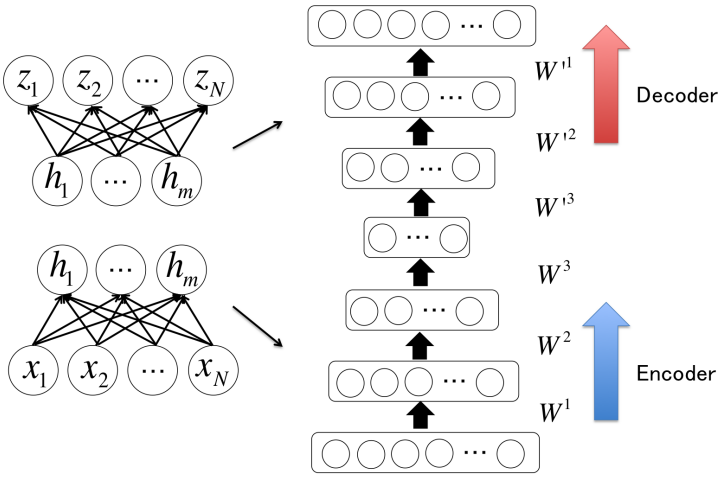

図 A.2: Deep Auto Encoder

付 録B　Restricted Boltzman Machine

Restricted Boltzman Machine とは，全結合ネットワークを利用するボルツマンマシンを，可視層と隠れ層の間での結合のみに制約することで，図 B.1 のような層状のネットワークとして実現したものである．

学習則は基本的にボルツマンマシンと同様である．従ってまず，ボルツマンマシンについて概説する．

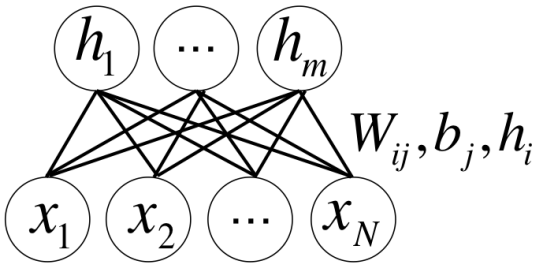

図 B.1: RBM

B.1　ボルツマンマシン

ボルツマンマシン [1] とは，0 か 1 かの 2 通りの値を持つ複数の素子（ニューロン）と，それらをつなぐネットワーク（シナプス）によって形成される．

素子 i の発火確率 $P(x_i := 1)$（$x_i = 1$ となる確率）は，素子への入力

$$E_i(\mathbf{x}) = \sum_{j=1}^{n} w_{ij} x_j + h_i \tag{B.1}$$

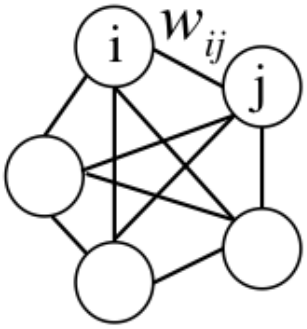

図 B.2: ボルツマンマシン

に応じて，次式で決定される．

$$P(x_i := 1) = \frac{1}{1 + exp(-E_i(\mathbf{x})/T)} \tag{B.2}$$

ここで，x_i は各素子の状態を，w_{ij} は素子 i と j の間の結合の強さを，h_i は，素子 i が発火状態になる閾値を表す．ただし，$w_{ij} = w_{ji}$ とする．また，T はネットワークの温度とも呼べる正のパラメータであり，$T \to 0$ の極限では，入力によらず $P(x_i := 1) = \frac{1}{2}$ となり，$T \to \inf$ の極限では，$P(x_i := 1)$ は階段関数となる．これはつまり，高温域においてボルツマンマシンは乱雑な動きをし，低温域では決定論的な動きをすることを意味する．$T \to 0$ でこの更新を繰り返すと，エネルギー関数

$$E(\mathbf{x}) = -(\sum_{i<j} w_{ij}x_j + \sum_i h_i) \tag{B.3}$$

が単調に減少し平衡状態（局所解）に落ち着く．一方，$T \neq 0$ では，エネルギーが増大する方向にも確率的にネットワークが変化し得る．従って，巡回セールスマン問題のように，問題をこのボルツマンマシンのエネルギー関数の最小化問題に帰着させた場合などは，このパラメータをうまくコントロールしながら，局所解を避けてより大域的な解を求めることが行われる．

さて，ある温度 T でボルツマンマシンを長時間作動させた場合，その発火パターンの分布は初期状態によらず，ボルツマン分布に収束する．

$$P(\mathbf{x}) = \frac{1}{Z} exp(-E(\mathbf{x})/T) \tag{B.4}$$

$$Z = \sum_{\mathbf{x}} exp(-E(\mathbf{x})/T) \tag{B.5}$$

ここで，E は，式 B.3 て定義されたエネルギー関数で，Z は式 B.4 の右辺を確率とするための規格化定数である．

ボルツマンマシンにおける学習とは，この分布を，外界から与えられた任意の確率分布 Q に近づけていくプロセスであると考えることができる．例えば，2 値のピクセルデータによって表現された N 個の手書き文字パターン $\{\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(N)\} \subset \{0,1\}^N$ のようなものをボルツマンマシンに記憶させる学習を想定すると，ボルツマンマシンの中に可視ノードを設定し，そのノードに画像のピクセルデータを割り当て，その複数の画像によって構成されるデータの確率分布

$$Q(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}(i) \tag{B.6}$$

を，ボルツマンマシンの平衡分布 P として再現することが考えられる．

このような学習の指標として，分布間距離 KL-divergence の最小化が導入される．

$$KL(Q||P) = \sum_{\mathbf{x}} Q(\mathbf{x}) log \frac{Q(\mathbf{x})}{P(\mathbf{x})} \tag{B.7}$$

具体的な更新則の導出を，RBM を例にとって次節にて行う．

B.2 制約付きボルツマンマシン

RBM は，前述したように，ボルツマンマシンの結合の一部を制限したネットワーク構造になっているが，学習則は変わらず KL-divergence の最小化である．ここまでの説明によって，更新則導出に必要な条件がそろった．以下にそれをまとめる．

$$P(x) = \frac{1}{Z} exp(-E(x)/T) \tag{B.8}$$

$$Z = \sum_{x} exp(-E(x)/T) \tag{B.9}$$

$$E(x) = -(\sum_{i<j} w_{ij} x_i x_j + \sum_{i} h_i x_i + \sum_{j} b_j h_j) \tag{B.10}$$

$$
\begin{aligned}
KL(Q||P_\theta) &= \sum_x Q(x) log \frac{Q(x)}{P(x)} \\
&= \sum_x Q(x) log Q(x) - \sum_x Q(x) log P(x)
\end{aligned}
$$

この KL-divergence の最小化を，最急降下法で解くことを考えると，重みの更新則を例にとって計算した場合，

$$
\begin{aligned}
\frac{\partial KL(Q||P)}{\partial w_{ij}} &= -\frac{\partial}{\partial w_{ij}} (\sum_x Q(x) log P(x)) \\
&= -\frac{\partial}{\partial w_{ij}} (\sum_x Q(x) log \{\frac{1}{Z} exp(-E(x)/T)\} \\
&= -\frac{\partial}{\partial w_{ij}} (\sum_x \{-Q(x)E(x)/T - Q(x) log Z\}) \\
&= -\frac{1}{T} \frac{\partial}{\partial w_{ij}} (\sum_x \{-Q(x)E(x) - Q(x) T log \{\sum_{x'} exp(-E(x')/T)\}\}) \\
&= \frac{1}{T} \{\sum_x Q(x) x_i x_j - \sum_x Q(x) \frac{\sum_{x'} x_i x_j exp(-E(x')/T)}{\sum_{x'} exp(-E(x')/T)}\} \\
&= \frac{1}{T} \{\sum_x Q(x) x_i x_j - \sum_x Q(x) \sum_{x'} x'_i x'_j P(x')\} \\
&= \frac{1}{T} \{\sum_x Q(x) x_i x_j - \sum_{x'_i} x'_i x'_j P(x')\}
\end{aligned}
$$

となる．w_{ij} をこの式に従って更新すれば損失関数が減少していく．(b_j，h_i の更新則も同様にして求められる．)

最後の式中の P を決定するには，ネットワークの平衡状態の分布を求めることが必要となる．具体的には，図 B.3 のように，ネットワークの上行・下降を繰り返してサンプリングを行うことで，平衡状態の分布を求めることになるが，これには，非常に長い時間がかかり，ボルツマンマシンを応用する上での問題点となっている．

ところが hinton らは，サンプリングを少数回で止めてしまう contrastive divergence 法を提案し，それがうまくいくことを実験的に示した [29]．本研究でも，1 回のみのサンプリングによって第 2 項を決定している．

以上のような方法で学習される RBM は，auto-encoder と同様にして DNN の pre-training に使用される．また，連続値入力データに対応したモデルも提案され

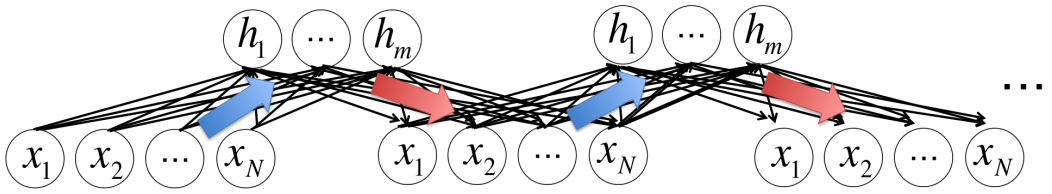

図 B.3: RBM におけるサンプリング

ており，この場合入力素子 i の発火確率は，

$$P(x_i) = exp(-\frac{(x_i - (\sum_j^m W_{ij}h_j + h_i))}{2\sigma^2}) \tag{B.11}$$

と定義される．学習則は，σ の影響を除いて 2 値ニューロンの場合と同様である．

付 録C　Convolution Neural Networks

Convolution Neural Networks（以下，CNN）は，人間の脳の視覚野にみられる構造 [31] を元に提案されたモデルで，図 C.1 のようにニューロン間の結合が前層の空間的な局所領域のみに限定される構造を持つ．この局所領域をその領域を担当するニューロン毎に平行移動しながら空間全体を覆い尽くすわけだが，この時の結合重みパラメータ（畳み込みフィルタ：h_{pqk}）は画像中のすべての場所で共有される．

$$y_{i'j'k} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p,j+q} h_{pqk} \tag{C.1}$$

このような構造を多層に積み重ねたネットワークは，福島らによってネオコグニトロン [25] として提案されて以降，LeCun らによる Back Propagation での学習法の確立 [38] を経て，現在では画像認識分野における標準的な方法として確立された [70].

現在使用される CNN では一般に，前述の構造に加えて pooling と呼ばれる機構が組み込まれている．これは，図 C.2 にあるように，前層の需要野内の成分を一成分に写像する変換であり，例えば最大値の値を次層の値とする変換がよく使用される．

$$y_{i'j'} = \underset{\{(p,q)|0<p<h,0<q<h\}}{max} x_{i+p,j+q} \tag{C.2}$$

多層ニューラルネットワークとして以上のような機構を繰り返し行うことは，入力の平行移動に対する不変性を段階的に加えていることに対応する．直感的な説明としては，入力の解像度を少しずつ落としながら異なるスケールで隣接する特徴の共起をとり，識別に有効な情報を選択的に上層へ渡していくことを行っていると解釈できる [70].

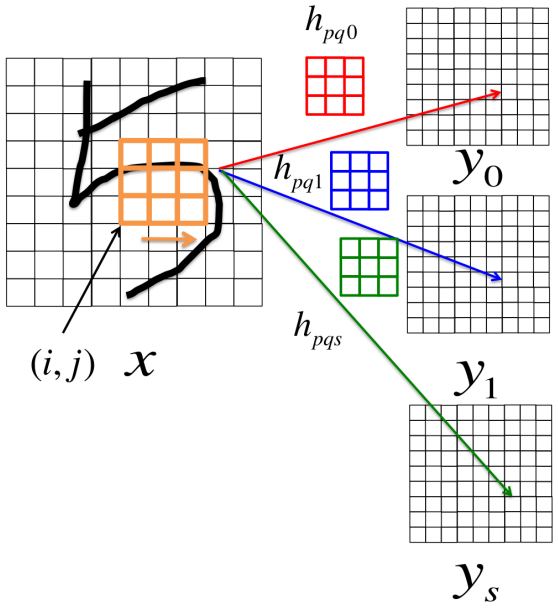

図 C.1: Convolution Network

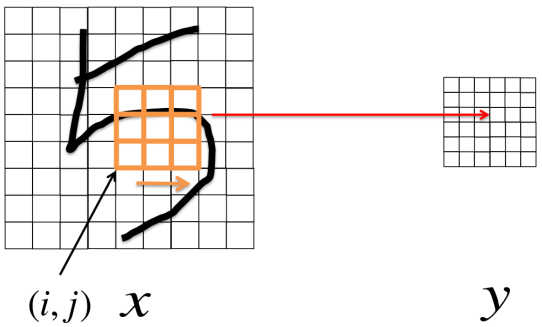

図 C.2: Pooling

付 録D　dropout

学習時に一定の確立pで中間層のユニットを無効化する手法[53]．極めて有効であることがわかっており，広く使用されている．識別時には全ての結合を復活させる．そのため，ユニットの出力をp倍して出力強度を調整する．

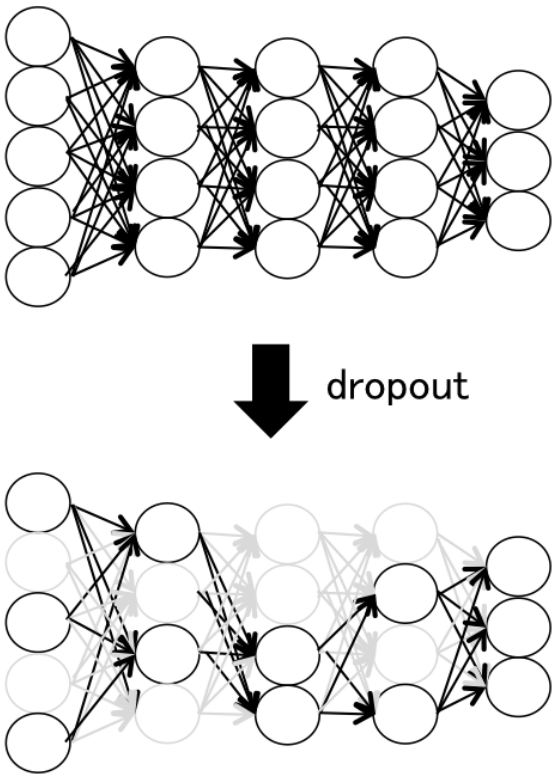

図 D.1: dropout

付 録E　ImageNetの右特異ベクトル一覧

実験５で得られた最大特異値と第２特異値に対応する右特異ベクトルを以下に掲載する．

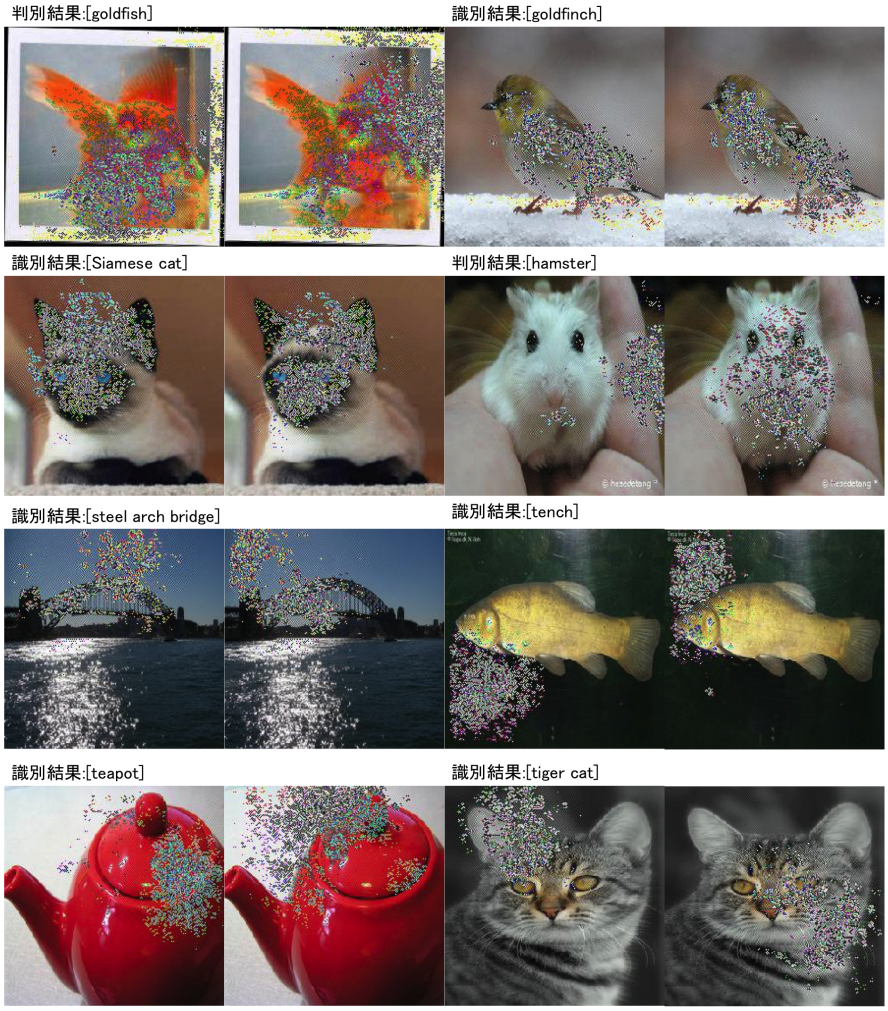

図 E.1: 右特異ベクトル１

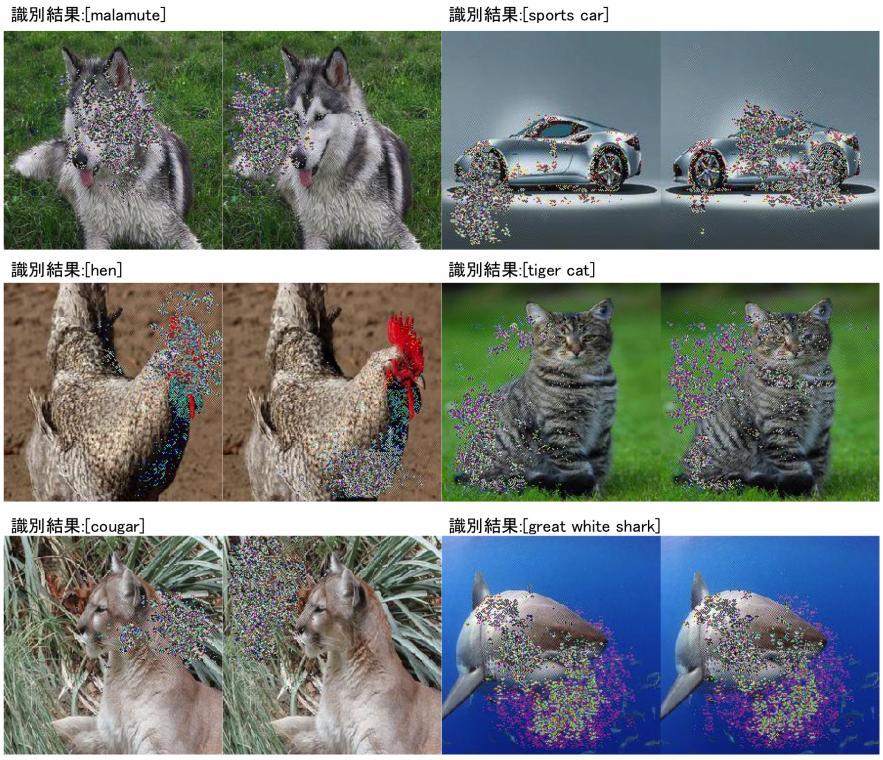

図 E.2: 右特異ベクトル 2