数理科学実践研究レター　2023–12　June 05, 2023

2 次元回転で不変・同変なタスクの学習

by

鶴崎 修功

UNIVERSITY OF TOKYO

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES

KOMABA, TOKYO, JAPAN

2 次元回転で不変・同変なタスクの学習

鶴崎修功[1]（東京大学大学院数理科学研究科）
Hisanori Tsurusaki (Graduate School of Mathematical Sciences, The University of Tokyo)

概 要

ニューラルネットワークの一種、多層パーセプトロンにおける学習では、線形変換層におけるパラメータの個数が少ないほど計算量が少なくなり、短時間で計算できる。データの対称性を考慮したニューラルネットワークを構成することで、学習の精度を保ちつつパラメータを減らすことができることが指摘されている。本研究では、2 次元データの回転に対して不変・同変な性質をもつ機械学習のタスクについて、ニューラルネットの層において不変・同変性を考慮し、そうしたときの線形変換層の次元を計算した。

1 はじめに

ニューラルネットワークの一種、多層パーセプトロン (MLP) における計算では、線形変換層と活性化層を繰り返す。線形変換層におけるパラメータの個数が少ないほど計算量が少なくなり、計算時間の短縮が期待できるが、精度が低下するリスクがある。
機械学習のタスクには、データが対称性をもつものがある。[DS17] では、データの置換に対して不変なタスク、すなわち入力データを置換しても結果が変わらないことが期待されるタスクと、置換で同変なタスク、すなわち入力データを置換すると結果が同様に置換されることが期待されるタスクを考え、それに対し、置換不変性・同変性を考慮したニューラルネットワークを構成する理論を与えた。
また、[IE19] では、[DS17] をうけ、グラフネットワーク上での置換同変・不変な学習タスクについて、MLP の線形変換層に必要なパラメータの個数、すなわち次元を具体的に計算した。
本研究では、対称性を置換から 2 次元の回転に変更し、2 次元データに巡回群 $\mathbb{Z}/m\mathbb{Z}$ の作用があって、この作用で不変・同変なタスクがあるとき、MLP の線形変換で探索すべき次元がどれくらい削減できるかを計算する。

2 データの回転不変性・同変性

定義 1 あるタスクがデータへの群作用によって不変であるとは、データに群を作用させる前と後で、期待される出力が変わらないことである。

定義 2 あるタスクがデータへの群作用によって同変であるとは、データに群を作用させて得た出力が、元のデータから得た出力に群を作用させて得られるものと等しくなることである。

例 3 猫が写った画像か、犬が写った画像が、回転した状態で与えられたときに、画像から写っているのが犬か猫か判定するタスクは、*2* 次元データの回転作用によって不変である。
また、画像が回転した状態で与えられたときに、もともと正しい向きはどの方向であったかを判定するタスクは、*2* 次元データの回転作用によって同変である。

3 行列による定式化

本研究では、群作用として 2 次元データの回転を考える。以下、タスクにおいて、データは実数値もしくは実数値の（ハイパー）グラフネットワーク、すなわちデータ要素のテンソル代数であるとし、

[1] htsuru@ms.u-tokyo.ac.jp

データ要素の個数は n であるとする。ただし、無限群 $SO(2)$ を扱うのは難しいため、回転作用の近似として、巡回群 $\mathbb{Z}/m\mathbb{Z}$ の作用が次のようなものであるとする。$m \times m$ 行列 B を

$$B = \begin{pmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 \\ 1 & & & \end{pmatrix}$$

とし、$n = md$ として、$n \times n$ 行列 A を

$$A = \begin{pmatrix} B & & \\ & \ddots & \\ & & B \end{pmatrix}$$

とおく。A は B を d 個対角に並べた形となる。$n = md$ 個のデータ要素からなる $n \times 1$ 行列

$$\begin{pmatrix} p_{11} \\ \vdots \\ p_{1m} \\ p_{21} \\ \vdots \\ \vdots \\ \vdots \\ p_{dm} \end{pmatrix}$$

に対して、$k \in \mathbb{Z}/m\mathbb{Z}$ の作用が A^k で表されるものになるとする。A は、p_{12} を 1 行目に、p_{13} を 2 行目に、...、p_{1m} を $m-1$ 行目に、p_{11} を m 行目に移す。
この作用は、データの 2 次元回転のうち、角度 $\frac{2\pi}{m}$ の回転のみを考慮した場合にあたる。

4　1 層 MLP における回転を考慮した探索次元の削減

MLP としてまず 1 層だけのものを考えることにする。

4.1　不変の場合

先に次の記号を定義する。

定義 4 $N \times M$ 行列 $X = (x_{ij})$ に対し、$NM \times 1$ 行列 $\mathrm{vec}(X)$ を

$$\mathrm{vec}(X) = \begin{pmatrix} x_{11} \\ \vdots \\ x_{N1} \\ x_{12} \\ \vdots \\ x_{N2} \\ \vdots \\ \vdots \\ x_{NM} \end{pmatrix}$$

で定める。

次の補題が計算によってわかる (cf. [IE19])。

補題 5 行列 X, Y, Z に対して $\mathrm{vec}(XZY) = Y^\top \otimes X \,\mathrm{vec}(Z)$ が成り立つ。

まずデータが実数値の場合を考える。回転作用を考慮しない場合、不変なタスクでは線形変換層は $n \times 1$ 行列 L_i で表され、L_i の各成分を探索することになる。すなわち、探索すべき次元は n となる。データが回転作用に関して不変であるということは、$L_i A^k = L_i$ $(0 \leq k < m)$ ということである。これは補題 5 より ${A^k}^\top \mathrm{vec}(L_i) = \mathrm{vec}(L_i)$ $(0 \leq k < m)$ と書き直せ、${A^k}^\top = A^{m-k}$ より、これは $A^k \mathrm{vec}(L_i) = \mathrm{vec}(L_i)$ $(0 \leq k < m)$ と同値である。$L_i = (l_{11}, \ldots, l_{1m}, l_{21}, \ldots, l_{dm})^\top$ とおき、これを解くと

$$\begin{cases} l_{11} = l_{12} = \cdots = l_{1m} \\ l_{21} = l_{22} = \cdots = l_{2m} \\ \qquad \vdots \\ l_{d1} = l_{d2} = \cdots = l_{dm} \end{cases}$$

が得られる。これにより、探索すべき次元は d となることがわかる。これは元の次元の $\frac{1}{m}$ である。
次に、データがグラフネットワークの形をしている場合を考える。データの 1 次テンソルの部分は、実数値の場合と同様に探索次元は $\frac{1}{m}$ となる。2 次テンソルの部分に対して、線形変換層は $n^2 \times 1$ 行列 L_{i2} で表される。このとき、データの回転作用で不変であるということは $L_{i2}(A^k \otimes A^k) = L_{i2}$ $(0 \leq k < m)$ を意味し、上と同様に考えるとこれは $A^k \otimes A^k \mathrm{vec}(L_{i2}) = \mathrm{vec}(L_{i2})$ $(0 \leq k < m)$ と同値となる。これを解くと同様に L_{i2} の成分が m 個ずつ等しくなることがわかる。すなわち、2 次テンソルの部分に対する線形変換層の探索の次元も $\frac{1}{m}$ となることがわかる。同様に、l 次テンソルの部分を考え、不変性を考慮した場合、その部分に対する線形変換層の探索の次元は、不変性を考慮しない場合の $\frac{1}{m}$ となる。したがって、データ全体を考えたとき、探索すべきパラメータの次元を $\frac{1}{m}$ に削減することができる。
以上は、巡回群 $\mathbb{Z}/m\mathbb{Z}$ の右正則表現を V をおいたとき、任意の正の整数 d, l に対して $(V^{\oplus d})^{\otimes l}$ の表現空間の不変部分空間の次元が $(V^{\oplus d})^{\otimes l}$ の表現空間の次元の $\frac{1}{m}$ であるという内容にいいかえることができる。

4.2 同変の場合

まずデータが実数値であるときを考えると、同変なタスクでは、MLP の線形変換層は $n \times n$ 行列 L_e で表され、探索すべき次元は n^2 となる。同変性の条件は $L_e A^k = A^k L_e$ $(0 \leq k < m)$ で表される。${A^k}^\top = A^{m-k} = {A^k}^{-1}$ より、これは ${A^k}^\top L_e A^k = L_e$ $(0 \leq k < m)$ と書き直せて、補題 5 よりこれは ${A^k}^\top \otimes {A^k}^\top \mathrm{vec}(L_e) = \mathrm{vec}(L_e)$ $(0 \leq k < m)$ と同値になる。これは $A^k \otimes A^k \mathrm{vec}(L_e) = \mathrm{vec}(L_e)$ $(0 \leq k < m)$ と書き直せて、不変の場合の議論より、この場合も L_e の成分が m 個ずつ等しくなることがわかる。よって、パラメータ探索の次元を $\frac{1}{m}$ に削減できる。
データがグラフネットワークの形であるときも、同様の議論で探索の次元を $\frac{1}{m}$ に削減できる。

5 不変・同変性を考慮した多層 MLP

不変な多層 MLP を構成するには、同変な層を重ね、最終層だけを不変な層にすればよい。同変な多層 MLP は、単純に同変な層を重ねればよい。どちらの場合でも、パラメータ探索の次元は $\frac{1}{m}$ になることがわかった。

6 終わりに

本論文では、巡回群 $\mathbb{Z}/m\mathbb{Z}$ による回転作用に対する対称性をもつタスクに対して、MLP のパラメータ探索を $\frac{1}{m}$ にできることを計算した。
実際のタスクで、例 3 のような画像を処理する場合を考えると、画像は格子状に並んだ画素の集合であり、回転によって閉じているとは限らない。このような場合、周りの画素から補間することで回転後のデータを生成する必要があると考えられる。このような場合に本研究の手法が有力であるかどうかを調査することは、今後の興味深い問題である。

7 謝辞

本研究で扱った興味深い課題を提供くださり、また多くの議論をしていただいた、株式会社ニコンの中村ちから様、信田 萌伽様、小池 哲也様にお礼を申し上げたいと思います。また、セミナーを主導してくださった東京大学の鮑園園様、同じセミナーで取り組んだ、東京大学、FMSP コース生の後藤祐輝様、Hu Xin 様にも感謝申し上げます。

参考文献

[DS17] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, Alexander Smola, *Deep Sets*, NIPS, 2017, arXiv:1703.06114

[IE19] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, Yaron Lipman, *Invariant and Equivariant Graph Networks*, ICLR, 2019, arXiv:1812.09902