物理学とディープラーニング (ゼミ)

B4 柳瀬調知

2023 年 4 月 21 日

第 4 章 4.2.2 節

・表現学習

　機械学習で写真を扱いたい場合、写真のピクセル値などを数値配列をしてまとめてデータ化し、さらに人間が分析に不要な情報を落とすなど、分析の精度を上げるための努力をする。機械学習に用いるデータの表現形態を”表現”というが、機械学習の性能のかなりの部分が問題に適した表現が作れるのかで決まってしまっていた。

　そこで登場したのが「表現学習」という考え方である。これは、問題に適した表現を機械に自発的に獲得させようというアプローチである。これまでは入力 x をそのまま回帰にかけていた $(\boldsymbol{x} \to y(\boldsymbol{x}))$ が、入力を問題に最適な表現に変換する方法も学習しつつ、その表現を回帰にかけるというイメージである。

$$\boldsymbol{x} \to \boldsymbol{h}(\boldsymbol{x}) \to y(\boldsymbol{h}(\boldsymbol{x})) \tag{1}$$

ニューラルネットワークが重要視されているのは、さまざまな状況で表現 $\boldsymbol{h}(\boldsymbol{x})$ を学習するのに、汎用的で強力な手法を提供してくれるから。

・順伝播型ニューラルネットワーク

　ディープ Boltzmann マシンでは、データを生成する確率モデル (生成モデル) をネットワークで表現することが目的だった。ここからは、$y = f(x)$ の変換にもネットワーク構造を導入する。

　図 4.2 にあるユニットを層状につないだもの。入力層、隠れ層、出力層からなる。以下では n 層のニューラルネットを考える。

・p120 と 121 の内容

　フーリエ級数展開では、基底関数の重ね合わせで周期関数を表現した。これに対し、ニューラルネットでは、シグモイド関数のような簡単な関数をネットワーク構造により重ね合わせることで、複雑な関数を表現する。(p120)

　ニューラルネットでは、関数に含まれるパラメータをファインチューニングしたり、引数の合成を行列で書けないようなものに拡張するのではなく、ネットワークを複雑にすることで表現力を向上させる。(p121)

・図 4.11 の 3 つの層（インプット層、隠れ層、アウトプット層）を詳しくみていく

(i) インプット層

入力ベクトル $\boldsymbol{x}$ の各成分を出力する。活性化関数をもたず、単に値 x_i を出すだけ。

$$x_{(0)i} = x_i \tag{2}$$

(ii) 隠れ層

第 l 層の i 番目のニューロンを考える。$(1 \leq l < n)$

このニューロンは第 $(l-1)$ 層からの $x_{(l-1)j} \quad (j = 1, 2, \dots)$ を受け入れるので、入力は $\sum_j W_{(l)i,j} x_{(l-1)j} + b_{(l)i}$ となる。ここで、$W_{(l)i,j}$ は第 $(l-1)$ 層の j 番目のニューロンと第 l 層の i 番目のニューロンの結合の重みである。

このニューロンの出力は、入力に活性化関数を作用させたもので、

$$x_{(l)i} = \varphi_{(l)} \left(\sum_j W_{(l)i,j} x_{(l-1)j} + b_{(l)i} \right) \tag{3}$$

となる。一般には、層ごとに共通の活性化関数を使うが、層が異なれば、異なる関数にしても構わない。

(iii) アウトプット層

最後の第 n 層が出力層に相当する。$\boldsymbol{x}_{(n-1)}$ が入力 $\boldsymbol{x}$ の”表現”であり、この表現 $\boldsymbol{x}_{(n-1)}$ の回帰分析をすることで、$\boldsymbol{x}$ と $\boldsymbol{y}$ の関係を推定するのが出力層の役割である。

$$y_i = x_{(n)i} = \varphi_{(n)} \left(\sum_j W_{(n)i,j} x_{(n-1)j} + b_{(n)i} \right) \tag{4}$$

第 $(n-1)$ 層までの部分は層ごとに高次な表現 $\boldsymbol{h} = \boldsymbol{x}_{(n-1)}$ を構成する役割を果たしており、出力層の第 n 層はこの表現 $\boldsymbol{h}$ を使って通常の機械学習を行っている。

・なぜ深層とするのか

図 4.4 のニューラルネットワークは隠れ層が 1 層だけのものになっている。普遍性定理によれば、1 層でも十分だといえる。$\epsilon \to 0$ の漸近的な振る舞いを「浅い」ものと「深い」もので比較すると次のようになる。

表 1　オーダー記法

	隠れ層の数	隠れニューロンの数
浅い	$< C\lvert\log\frac{1}{\epsilon}\rvert$	$\geq C\lvert(\frac{1}{\epsilon})^n\rvert$
深い	$\simeq C\lvert\log\frac{1}{\epsilon}\rvert$	$\leq C\lvert\left(\log\frac{1}{\epsilon}\right)^n\rvert$

・隠れニューロン数 n_h を等しいとしたとき、1 層と 2 層でのパラメータ数の比較

層を増やすと、パラメータの数の増え方も急激に増えていくので、隠れ層をむやみに増やせばいいわけではない。