

物理学とディープラーニング (ゼミ)

B4 柳瀬調知

2023 年 4 月 12 日

第 1 章

・「学習」とは何か

学習とは、何か目的を定めて最適化すること。もう少し数学的に言えば、正確な分類 (回帰) をするために数理的な学習モデルに含まれるパラメータを最適化すること。

学習モデルを考えると、従来は異なる問題ごとに特化した最適なモデルが模索されていた。そのため、ある程度の人間の勘や知識が要求されていた。しかし、人間の脳は個別の問題ごとに別の回路を作っているわけではなく、もっと汎用性のある構造になっているはずだ。そこで、人間の脳に似せたモデルが設計出来たら汎用的な枠組みであらゆる問題を扱えるのではないかとということで考案されたのがディープラーニング。

しかし、闇雲にディープラーニングすればよいというわけではなく、従来の回帰分析やベイズ推定とも使い分けることが重要。p7 にあるようにディープラーニングでも過学習を起こす。

・物理との関係

物理は古典力学にしても量子力学にしても「最適化」の問題を解くプロセスなので、機械学習と類似している (p8)。特に Boltzmann マシンは統計力学との関係が深い。このことから、物理を知っておくことは重要だ。

第 2 章

学習とは、 $y = f(x)$ の $f(x)$ の関数形を最適化すること。

y が離散的なら分類、連続的なら回帰。

1. 教師あり学習

与えられたデータ $(x^{(\mu)}, y^{(\mu)})$ を再現するように $f(x^{(\mu)}; \theta)$ を最適化すること。具体的な 2 次関数の例を挙げる (図 2.2 参照)。教師あり学習の特徴はインプット $x^{(\mu)}$ に対する「答え」の $y^{(\mu)}$ が与えられていること。

一般に訓練データに対して正解と予測値とのずれは $\{y^{(\mu)} - f(x^{(\mu)}; \theta)\}$ に依存しており、またこのずれは損失関数

$L(\{y^{(\mu)} - f(x^{(\mu)}; \theta)\})$ と呼ばれている。「損失関数」を使うと、教師あり学習とは、訓練データを再現するように θ を調整し損失関数 $L(\{y^{(\mu)} - f(x^{(\mu)}; \theta)\})$ を最小化することだと説明できる。

応用例：手書き文字の判別

2. 強化学習

教師あり学習と違って、訓練データで答えとなる y は与えられていない。

「環境」や「エージェント」といった言葉のほかに状態 s 、行動 a 、報酬 r 、価値 (Q 値) といった概念が登場する。

・具体的な自転車の例 (p19)

エージェント：自転車の乗り手、環境：路面状況など

自転車の走行距離を最大化するという目的が (Q 値) である (最大化したい量)。各ステップで自転車を安定化させる行動 a をとり、その結果自転車は状態 s になる。この状態 s に応じて報酬 $r(s,a)$ がもらえる。次のステップでは前のステップにおける状態に応じた行動をとり、あとはそのくり返しである。 Q 値はある時刻における値ではなく積算値であるため、安定化する行動 (報酬の高い行動) をとり続けた結果、 Q 値が最大化される。

以上のことを式で表したのが (2-1) であると考えられる。ただし、 γ は無限和を収束させるための因子。

3. 教師なし学習

特徴量とは、インプット $\{x^{(\mu)}\}$ のこと。特徴量の張る空間を特徴空間という。

・教師なし学習のアルゴリズムの一例「オートエンコーダー」:図 2.5 参照

x と \hat{x} がなるべく近くなるように $f(x)$ と $g(y)$ の関数形を決める。これは、損失関数 $L(\{x^{(\mu)}, \{\hat{x}^{(\mu)}\})$ が最小となる θ を決めることに相当する。

教師あり学習とは違って、訓練データが $\{x^{(\mu)}\}$ のみである。それと、2つの出力はどちらも重要。

応用例：異常検知

学習後に $\{x^{(p)}\}$ と $\{\hat{x}^{(p)}\}$ が大きく異なる。このとき、 p 番目は異常。

4. 転移学習 (図 2.7 を参照)

問題設定を少し変えただけなら、学習モデルを一部変更するだけでいいかもしれない。

図 2.7 で f_1 が特徴抽出、 f_2 が分類を担当しているとする。例えば、動物の画像分類から車や船の画像分類に目的が変更されても、 f_1 の方は最適化したパラメータ θ_1 をそのまま使えるかもしれないから、 f_2 のみを最適化すればよさそう。

・ファインチューニング

上とほぼ同じだが、 f_1 を変更前の学習モデルからスタートすることで微調整だけで最適なパラメータ θ_1 が見つかるかもしれないという方法もある。こうすることで、学習にかかるコストを削減できる。