ISSN 2186-5647

OS-7

統計力学で機械学習を理解する

日大生産工　○原　一之

1.　まえがき

近年の人工知能の発展はめざましく、人を遥かに超える性能を示すようになっている。この背景には、半導体技術、コンピュータ技術の発展により、コンピュータの性能が飛躍的に向上し、従来から開発されていたアルゴリズムがその能力を発揮できるようになったこと、ビックデータの出現により、質のよりデータが大量に入手できるようになったことなどがある。その一方で、ブラックボックスとして扱われる機械学習に対する不安・不満も一部では顕在化している。それ対して可視化により機械学習がどのように情報処理するかを理解しようとする研究も現れている。本報告では，オンライン学習という機械学習の一種の学習過程を微視的な状態とみなし、学習後の性能を解析する統計力学を用いて解析する方法を紹介する。

2.　統計力学とモデルの定式化

統計力学の目的については、久保亮五先生が著書[1]で「(1)巨視的物理学において、経験的に定めるべきものとされていた巨視的現象の特徴を微視的法則をもとにして、その構造から導き出すこと、(2)一般に、巨視的な物理法則そのものを、微視的法則から導き出すこと」と書かれている。統計力学による機械学習の解析はこの目的を方針とする。

本報告では機械学習としてオンライン学習を扱う。オンライン学習とはデータを受け取ると即座にそのデータで学習をし、モデルパラメータを更新する学習法の総称である。ここでモデルパラメータとは学習機械の挙動を決定するパラメータである。オンライン学習における微視的法則は一個の学習データに基づいてモデルパラメータを変更するための学習方程式（式(1)）である。学習機械のモデルとして単純パーセプロトンを用いる。この学習モデルは最も単純な学習機械であり、解析が明快である。

$$y=\sum_{i=0}^{n} x_i w_i \text{。} \tag{1}$$

ここでyは出力、x_iはi番目の入力、w_iはi番目の入力に対する荷重である。このモデルは式(1)をデータ$\mathbf{x}$に対するyを予測する線形回帰問題、式(1)を平面を$y<0$と$y>0$の２つの領域に分離するdecision boundaryとして線形分離問題と考えることができる。

次に式(1)を用いて学習方法を定式化する。簡単のため、問題は線形回帰とする。データ$\mathbf{x}$はN次元ベクトルとし、各要素x_iは正規乱数N(0,1)で生成する。学習には教師・生徒の定式化を用いる。この定式化では教師が生徒の入力に対する所望の出力を与える。教師出力tは生徒と同じ構造の学習機械のモデル（式(1)）で生成すると仮定する。教師のモデルパラメータをN次元ベクトルw^tとし、各要素w_i^tは正規分布N(0,1/N)で生成する。同様に生徒の初期モデルパラメータの各要素$w_i^s(0)$も正規分布N(0,1/N)で生成する。つぎに誤差を定義する。誤差は生徒と教師に同じ入力$\mathbf{x}$を入力時の出力の２乗誤差（式(2)）で定義する。

$$\epsilon=\frac{1}{2}(y-t)^2 \tag{2}$$

２乗誤差は入力$\mathbf{x}$によって式(2)の値が変わる可能性があるため、$\mathbf{x}$に依存しない誤差として汎化誤差を導入する。汎化誤差は式(2)を無限個の入力$\mathbf{x}$に対する平均と定式化する。式(3)の<・>は無限大個の入力$\mathbf{x}$に対する平均を表す。

$$\epsilon_g=\langle\frac{1}{2}(y-t)^2\rangle \tag{3}$$

学習は式(3)を減少するように生徒のモデルパラメータ$w_i^s(\mathrm{m})$を更新することで行う。ここでmは学習回数を表す。モデルパラメータの更新則として確率的勾配法を用いた。確率的勾配法では式(3)の勾配を、ϵのモデルパラメータ$w_i^s(\mathrm{m})$の偏微分で求める。式(4)に更新式を示す。

$$w_i^s(\mathrm{m}+1)=w_i^s(\mathrm{m})+\frac{\eta}{N}\big(t(m)-y(m)\big)x_i(m) \tag{4}$$

ここまでの計算は微視的な量である。式(4)に従ってモデルパラメータを更新した時の式(3)を求めることで学習の評価ができる。ここで式(3)は巨視的な量であることに注意する。我々は式(3)の計算に統計力学を用いる。

3.　統計力学による解析

式(3)の平均を計算すると次式が得られる。

Understanding behavior of machine learning using statistical mechanics method

Kazuyuki HARA

$$\epsilon_g = \frac{1}{2}(1 - 2R + Q) \tag{5}$$

ここでRはモデルパラメータ$\boldsymbol{w^t}$と$\boldsymbol{w^s}$との内積である。Qは生徒のモデルパラメータ$\boldsymbol{w^s}$のノルムである。RとQは巨視的な変数であり、オーダパラメータと呼ばれる。式(5)を求める過程で$\langle t^2\rangle$、$\langle ty\rangle$、$\langle y^2\rangle$の計算が必要であり、統計力学が使われる。ここではモデルの次元Nが無限大の極限を想定し、中心極限定理と大数の法則が成立するとした。通常は我々が知ることのできない教師のモデルパラメータを明示的に用いている点に注意する。教師・生徒の定式化ではこのような枠組みを用いることにより、通常は知り得ないモデルパラメータw^sの評価を行なう。つまり教師と生徒のモデルパラメータのなす角をRで、ノルムの違いをQで直接評価することができる。

Rの時間発展は、式(4)に$R = \boldsymbol{w^t} \cdot \boldsymbol{w^s}$を用いると

$$\boldsymbol{w^s}(m+1) \cdot \boldsymbol{w^t} = \boldsymbol{w^s}(\mathrm{m}) \cdot \boldsymbol{w^t} + \frac{\eta}{N}\langle (t(m) - y(m))\boldsymbol{x}\rangle \tag{6}$$

となる。$t = {}^{m}/_{N}$とし、N無限大の極限をとり、両辺の入力平均を計算すると、Rの微分方程式が導出できる。

$$\frac{dR}{dt} = \eta(1 - R) \tag{7}$$

ここではRの自己平均性を仮定した。またQの時間発展は式(4)よりQ=$\boldsymbol{w^s} \cdot \boldsymbol{w^s}$なので式(4)の両辺を２乗して

$$\boldsymbol{w^s}(m+1) \cdot \boldsymbol{w^s}(m+1) = \boldsymbol{w^s}(\mathrm{m}) \cdot \boldsymbol{w^s}(m) + 2\frac{\eta}{N}\langle (t(m) - y(m))\boldsymbol{x} \cdot \boldsymbol{w^s}(\boldsymbol{m})\rangle + \frac{\eta^2}{N^2}(t(m) - y(m))^2 \boldsymbol{x} \cdot \boldsymbol{x} \tag{8}$$

となる。Rの微分方程式と同様の手続きでQの微分方程式を導出できる。

$$\frac{dQ}{dt} = 2\eta(R - Q) + \eta^2(1 - 2R + Q) \tag{9}$$

式(7)と式(9)より、微分方程式が停止した時にR=1、R=Q、$1 - 2R + Q = 0$が成立することがわかる。ここで$1 - 2R + Q = 0$が成立すると式(5)より汎化誤差が０になることに注意する。つまり式(4)を用いて学習を行うと、R=1つまり教師と生徒のモデルパラメータのなす角は０となり、それぞれのノルムが一致すること、R=Qより、Q=1になり、教師のモデルパラメータのノルム（定義より$|w^t| = 1$）と等しくなることがわかる。

このように、教師・生徒の定式化を用いているものの、学習の挙動を詳細に理解できる点が統計力学を用いた解析の利点である。

次に式(7)および(9)を解析的に解く。

$$\mathrm{R(t)}=1-(1-\mathrm{R}(0))e^{-\eta t} \tag{10}$$

$$\mathrm{Q(t)}=1-2(1-\mathrm{R}(0))e^{-\eta t}+(\mathrm{Q}(0)-2\mathrm{R}(0)+1)e^{-\eta(2-\eta)t} \tag{11}$$

式(11)より、右辺第２項の指数関数の肩の数式$(-\eta(2-\eta))$の微分が０になる条件よりη=1の時に学習が最も早く収束し、第２項の指数関数の肩の数式が負である条件より$\eta \geq 2$でQが指数関数的に増大し、誤差も増大する。

4. 実験結果および検討

統計力学を用いて導出した式(5)、(10)、(11)と計算機シミュレーションの結果を比較することにより解析結果の正当性を検証する。計算機シミュレーションでは式(4)を用いてモデルパラメータを更新する。実験条件は次の様にした。システムサイズN=100、R(0)=0、Q(0)=0、ηは0.2、1.0、1.9、2.0とした。

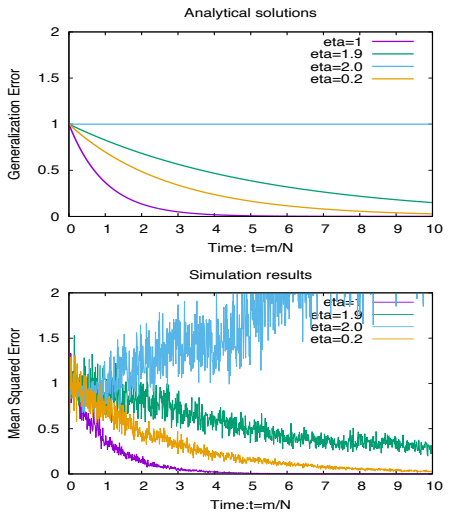

図1：解析解とシミュレーション結果の比較

図1に結果を示す。上が解析解、したがシミュレーション結果である。両者が一致したことから解析解の正当性が示された。

5. まとめ

本稿では統計力学の手法を用いて機械学習の挙動を理解する方法を紹介した。その結果、学習方程式を共通に使用するシミュレーションと解析解が一致することを示した。また、解析解により機械学習の挙動を理解できる。

参考文献

[1] 久保亮五, “統計力学”, 共立出版（1952）p.5.
[2] 西森秀稔, “スピングラス理論と情報統計力学”（1999）.