

パーセプトロンのノイズ有り学習

奈良女子大学 理学部 物理科学科 上江洲 達也

1. Introduction

現実の系には必ず何らかのノイズが存在するため、学習の問題においても、ノイズの効果を調べることは極めて意味深い。ここでは、教師パーセプトロンに外部ノイズが加わった時の生徒パーセプトロンによる学習の問題を、統計力学的手法を用いて解析する。

2. 一層パーセプトロンの学習

2.1. 問題設定

教師パーセプトロンの荷重ベクトル w^0 、生徒パーセプトロンの荷重ベクトル w 、例題ベクトル x をいずれも、 N 次元ベクトルとし、その大きさは $\sqrt{N}$ であるとする。 μ 番目の例題 x^μ に対する生徒の出力 r^μ は、

$$r^\mu = \text{sgn}(h^\mu), \quad h^\mu \equiv (x^\mu \cdot w)/\sqrt{N}$$

である。一方、 μ 番目の例題 x^μ に対して、教師が $r^{0\mu} = +1$ を出力する確率を次の形に仮定する。

$$p_r(+1|x^\mu) = P(h^{0\mu}) = \frac{1 + P(h^{0\mu})}{2}, \quad h^{0\mu} \equiv (x^\mu \cdot w^0)/\sqrt{N}.$$

学習アルゴリズムとしては、ここでは、Gibbs アルゴリズムを用いる。 p 個の例題とそれに対する教師出力の組、training set

$$\{x^\mu, R^\mu\}, \mu = 1, \dots, p,$$

が与えられたとき、生徒 w のエネルギー $E_p[w, x]$ を、 p 個の出力のうち生徒出力と教師出力が異なる回数と定義する。そして、生徒を $\exp(-\beta E_p[w, x])$ に比例する確率で選択する。 $\beta = 1/T$ で、 T は、「温度」。特に、 $T \rightarrow +0$ は、誤り最小アルゴリズム (Minimum error algorithm) となる。

学習の到達度を測るために、汎化誤差 ϵ_g を計算する。これは、新しい例題が与えられたときに、教師出力と異なる出力を出す確率の平均値である。

$$\epsilon_g = \langle P(u^0)(1 - \Theta(u)) + (1 - P(u^0))\Theta(u) \rangle_x.$$

汎化誤差の振舞いは、荷重ベクトルの種類や、ノイズの入り方等に依存している。ここでは、次のような場合を取り扱う。

○荷重ベクトルの種類。

・連続的な場合 (Spherical case)。各成分が連続変数。・離散的な場合 (Ising case)。 $w_i = \pm 1$ 。

○ノイズのタイプ

・入力ノイズ。入力される例題ベクトルにノイズが入る。平均 0、分散 σ のノイズ η_i が、各成分 x_i に独立にはいるとすると、例題 x にたいして、 $r^0 = +1$ となる確率 $p_r(+1|x)$ は、

$$p_r(+1|x) = H[-(x \cdot w^0)/(\sigma\sqrt{N})], \quad H(u) = \int_u^\infty dx \exp[-x^2/2]/\sqrt{2\pi}.$$

・出力ノイズ。教師出力が、確率 λ で、反転する。

$$p_r(+1|x) = \lambda + (1 - 2\lambda)\Theta[(x \cdot w^0)/\sqrt{N}]$$

・一般的に、次のようにかける。

$$p_r(+1|\mathbf{x}) = P(h^\circ) = \frac{1 + P(h^\circ)}{2}, \quad h^\circ \equiv (\mathbf{x} \cdot \mathbf{w}^\circ) / \sqrt{N}.$$

$P(u)$ の $u = 0$ 近傍での振舞いを $P(u) \simeq a \operatorname{sgn}(u)|u|^\delta$, とし、このクラスを $P_\delta(u)$ と表す。

入力ノイズは、 $\delta = 1$ であり、出力ノイズは $\delta = 0$ である。

このとき、汎化誤差は、

$$\begin{aligned} \epsilon_g &= \epsilon_{\min} + \int_0^\infty Dy H(y/\zeta) [P(y) - P(-y)], \\ \epsilon_{\min} &= \frac{1}{2} - \frac{1}{2} \int_0^\infty Dy [P(y) - P(-y)] = \int_0^\infty Dy (1 - P(y)) + \int_{-\infty}^0 Dy P(y), \end{aligned}$$

となる。ここで、 $Dy = \exp(-y^2/2)/\sqrt{2\pi}$, $\zeta = \frac{\sqrt{1-R^2}}{R}$. $\epsilon_{\min}$ は、 $\mathbf{w} = \mathbf{w}^\circ$ のときの、汎化誤差。 R は、教師と生徒のオーバーラップ。 ϵ_g は R の単調減少関数となる。

2.2. R の計算

レプリカ法を用いて R を計算する。我々は、RS 解、即ち、全ての量が、レプリカ指数 α, β に依存しない解と、この解が不安定な場合は、1RSB 解、即ち、レプリカ対称性を一回破る解を考える。具体的計算は、文献 [1, 2, 3] を参照。

2.3. 結果

鞍点方程式を解いて、 R を計算することにより、汎化誤差が求まる。結果は、ノイズのタイプや、荷重ベクトルの種類によって、2つに分類される。

$$\Delta\epsilon_g \equiv \epsilon_g - \epsilon_{\min} \propto \alpha^{-\gamma}, \quad \alpha \gg 1 \text{ のとき.}$$

$$\Delta\epsilon_g = 0 \text{ (完全学習), } \alpha > \alpha_p \text{ のとき.}$$

I. 一層、 $\mathbf{w}^\circ, \mathbf{w}$ 共に連続、関数族 P_δ

汎化誤差は、常に、ベキ的。

$$\Delta\epsilon_g = \alpha^{-\gamma}.$$

1. ギブスアルゴリズム及び 誤り最少アルゴリズム。

$\Delta\epsilon_g$ の漸近形

	RS 解	1RSB 解
出力ノイズ	α^{-1}	α^{-1}
入力ノイズ	$\alpha^{-\frac{1}{2}}$	$\alpha^{-\frac{2}{3}} (\ln \alpha)^{\frac{1}{3}}$
P_δ ($\delta > 0$)	$\alpha^{-\frac{1+\delta}{1+3\delta}}$	$\alpha^{-\frac{1+\delta}{1+2\delta}} (\ln \alpha)^{\frac{1+\delta}{2(1+2\delta)}}$

2. 最適温度の存在について。

・出力ノイズの場合

$$\Delta\epsilon_g \simeq \psi_0^{(RS)}(T) \alpha^{-1} : \text{RS 解}, \quad \Delta\epsilon_g \simeq \psi_0^{(1RSB)}(T) \alpha^{-1} : \text{1RSB 解}.$$

$\psi_0^{(RS)}(T)$ は、 T_0^* で最小値となり、このとき、RS 解は安定。1RSB 解を考慮しても最小である。

・入力ノイズの場合

$$\Delta\epsilon_g \simeq \{\psi_1^{(RS)}(T)\}^{\frac{2(1+\delta)}{1+3\delta}} \alpha^{-\frac{1+\delta}{1+3\delta}} : \text{RS 解},$$

$$\Delta\epsilon_g \simeq \psi_\delta^{(1RSB)}(T) \alpha^{-\frac{1+\delta}{1+2\delta}} (\ln \alpha)^{\frac{1+\delta}{2(1+2\delta)}} : \text{1RSB 解}.$$

RS 解は、

$$\psi_1^{(RS)}(T) \propto \sqrt{1/T} \text{ for } T \ll \alpha^{-1},$$

となるが、この解は、漸近的に不安定。一方、1RSB 解は、

$$\psi_\delta^{(1RSB)}(T) \propto \delta^{\frac{1+\delta}{2(1+2\delta)}} \text{ for } \alpha^{-\frac{\delta}{3(1+2\delta)}} \ll T \ll \alpha^{\frac{\delta}{1+2\delta}}$$

となり、広い温度領域で一定。

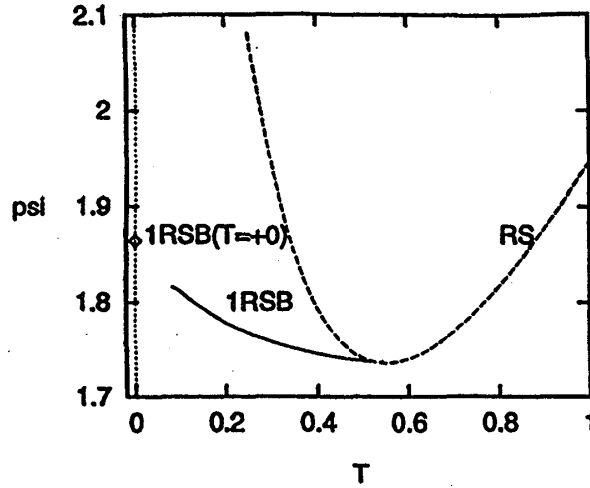


図 1: 出力ノイズの学習曲線の α^{-1} の係数の温度依存性。 $P(u) = 0.5 \operatorname{sgn}(u)$ 。点線と実線は、各々 $\psi_0^{(RS)}(T)$ と $\psi_0^{(1RSB)}(T)$ 。菱形は誤り最小アルゴリズムでの 1RSB 解の係数 $\psi_{MEA}^{(1RSB)}$ 。

II. 一層、 w^o, w 共に離散的、関数族 P_δ 、

$\Delta\epsilon_g$ の漸近形

RS 解は、漸近的に不安定。

	1RSB 解
$\delta > \frac{1}{2}$	$(\ln \alpha / \alpha)^{\frac{1+\delta}{2\delta-1}}$
$\delta = \frac{1}{2}$	$e^{-3F_0\alpha}, F_0$ は定数
$\delta < \frac{1}{2}$	完全学習

3. 二層パーセプトロンの学習 – non-overlapping case –

3.1. 問題設定

入力層 $N = MK$ 個、中間層 K 個、出力層 1 個のパーセプトロンを扱う [4]。教師パーセプトロンの荷重ベクトル w_l^t ($l = 1, \dots, K$), 生徒パーセプトロンの荷重ベクトル w_l^s ($l = 1, \dots, K$), 例題 x_l のいずれも、 M 次元ベクトルで、大きさは $\sqrt{M}$ とする。

μ 番目の例題 $\{x_l^\mu\}$ に対する教師出力 σ_t^μ は、ノイズ無しするとき、

$$\sigma_t^\mu = B_t(\{\sigma_l^\mu\}), \quad \sigma_l^\mu = \text{sgn}(h_{l,\mu}^t), \quad h_{l,\mu}^t \equiv (x_l^\mu \cdot w_l^t)/\sqrt{M}.$$

とする。ここで、 $B_t(\{\sigma_l^\mu\})$ は、教師を特徴づける Boolean function。教師信号にノイズが入るときは、教師が σ_t^μ を出力する確率を次のように表す。

$$p_r(\sigma_t^\mu | \{x_l^\mu\}) = \mathcal{P}(\sigma_t^\mu | \{h_{l,\mu}^t\}).$$

μ 番目の例題 $\{x_l^\mu\}$ に対する生徒の出力 σ_s^μ は、

$$\sigma_s^\mu = B_s(\{\sigma_l^\mu\}), \quad \sigma_l^\mu = \text{sgn}(h_{l,\mu}^s), \quad h_{l,\mu}^s \equiv (x_l^\mu \cdot w_l^s)/\sqrt{M}.$$

ここで、 $B_s(\{\sigma_l^\mu\})$ は、生徒の出力を特徴づける Boolean function。

3.2. モデルとその結果

レプリカ法を用いて、鞍点方程式を求め、 R を計算することにより、汎化誤差が求まる。

w^t, w^s は、共に連続とし、ノイズは、出力ノイズと入力ノイズを扱う。また、RS 解のみを考える。更に、 $B_t = B_s = B$ とし、permutation symmetry を仮定する。

結果

$\Delta\epsilon_g$ の漸近形

	RS 解
決定論的	α^{-1}
出力ノイズ	α^{-1}
入力ノイズ	$\alpha^{-1/2}$

参考文献

- [1] T. Uezu and Y. Kabashima, J. of Phys. **A29**(1996),L55: T. Uezu, Y. Kabashima, K. Nokura and N. Nakamura, J. of Phys. Soc. Japan **65**(1996),3797.
- [2] T. Uezu and Y. Kabashima, J. Phys. A: Math. Gen.**29**(1996),L439.
- [3] T. Uezu, J. Phys. A: Math. Gen.**30**(1997),L777.
- [4] B. Schottky, J. Phys. **A28**(1995),4515: B. Schottky and U. Krey, J. Phys. A: Math. Gen.**30**(1997),8541.