パーセプトロンの教師-生徒設定の統計力学

市川佑馬 *

2023 年 5 月 12 日

概要

教師と生徒の重みが超球上に制約された場合のパーセプトロンのレプリカ計算のメモ. 主に, [Engel and Van den Broeck, 2001] の計算を field computation の方法に修正したものである.

1 設定

まずは, データの性質, 学習アルゴリズムや汎化誤差についてまとめる.

1.1 データの性質

この問題では, 一次元出力からなる分類問題について考える.

$$\mathcal{D} = \{(\boldsymbol{x}^\mu, y^\mu)\}_{\mu=1}^{M}, \quad \boldsymbol{x}^\mu \in \mathbb{R}^N, \;\; y^\mu \in \{-1, +1\}.$$

さらに, この問題では, 具体的に $\boldsymbol{x}^\mu$, y^μ の生成ルールを次のように定める:

$$\begin{aligned} &\boldsymbol{x}^\mu \in \mathbb{R}^N, \;\; \boldsymbol{x}^\mu \sim \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{I}_N), \\ &y^\mu = \mathrm{sign}\left(\frac{1}{\sqrt{N}} \boldsymbol{w}_0^\top \boldsymbol{x}^\mu\right), \;\; \boldsymbol{w}_0 \sim \delta(\|\boldsymbol{w}_0\|^2 - N) \end{aligned}$$

ここで, $\boldsymbol{w}_0$ を教師ベクトルと呼ぶことにする.

1.2 学習手法

分類問題を解くための学習モデルとして, データの生成過程と同じモデルを仮定する.

$$\hat{y} = \mathrm{sign}\left(\frac{1}{\sqrt{N}} \boldsymbol{w}^\top \boldsymbol{x}\right),$$

ここで, 教師ベクトル $\boldsymbol{w}_0$ に対して, $\boldsymbol{w}$ を生徒ベクトルと呼ぶことにする. 目的関数は, ラベル y と予測 $\hat{y}$ の符号が異なる時のみ 1 となる以下の関数を用いる*1:

$$\mathcal{L}(\boldsymbol{w}; \mathcal{D}) = \sum_{\mu=1}^{P} \Theta\left(-\frac{\boldsymbol{w}_0^\top \boldsymbol{x}^\mu}{\sqrt{N}} \frac{\boldsymbol{w}^\top \boldsymbol{x}^\mu}{\sqrt{N}}\right) \tag{1}$$

この目的関数の最適化アルゴリズムとしては, Hebb 則や Perceptron 則が有名である.

* 東京大学大学院総合文化研究科広域科学専攻相関基礎科学系 福島研究室

*1 線形特徴量 $\boldsymbol{w}_0^\top \boldsymbol{x}$, $\boldsymbol{w}_0^\top \boldsymbol{x}$ の関数であれば, 後のレプリカ計算の一部分が変わるだけで, 同じような手順で解析することができる.

1.3 汎化誤差

汎化誤差は, 次のように定める:

$$\epsilon_g = \mathbb{E}_{\boldsymbol{x},\boldsymbol{w}_0}\left[\Theta\left(-\frac{\boldsymbol{w}_0^\top \boldsymbol{x}}{\sqrt{N}}\frac{\hat{\boldsymbol{w}}^\top \boldsymbol{x}}{\sqrt{N}}\right)\right], \tag{2}$$

ここで, 教師ベクトルにも生徒ベクトルと同様の規格化を行い

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}:\|\boldsymbol{w}\|^2=N}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{w};\mathcal{D}),$$

とする. $N \to \infty$ の場合, $u_0 = \boldsymbol{w}_0^\top \boldsymbol{x}/\sqrt{N}$, $u = \boldsymbol{w}^\top \boldsymbol{x}/\sqrt{N}$ は中心極限定理より次のように表せることに着目する:

$$\mathbb{P}[u_0, u] = \frac{1}{2\pi\sqrt{1-m^2}} e^{-\frac{u_0^2+u^2-2mu_0u}{2(1-m^2)}}, \quad m = \frac{1}{N}\boldsymbol{w}_0^\top \boldsymbol{w}.$$

ゆえに, 汎化誤差は m を用いて

$$\begin{aligned}
\epsilon_g &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{w}_0}\left[\Theta\left(-\frac{\boldsymbol{w}_0^\top \boldsymbol{x}}{\sqrt{N}}\frac{\hat{\boldsymbol{w}}^\top \boldsymbol{x}}{\sqrt{N}}\right)\right] \\
&= \int \mathbb{P}[u_0, u] du_0 du \Theta(-u_0 u) \\
&= \frac{1}{\pi}\arccos(m)
\end{aligned}$$

と表せる. この式の意味は, [西森, 2016] が詳しい.

1.4 Gardner 体積

やや天下り的だが, 次の量を考えると便利である.

$$V(\mathcal{D}) = \int d\boldsymbol{w}\delta(\|\boldsymbol{w}\|^2 - N)e^{-\beta\mathcal{L}(\mathcal{D})}$$

ここで, 最適化の制約条件は, $\delta(\|\boldsymbol{w}\|^2 - N)$ を用いて表した. この量は, Gardner 体積と呼ばれ, 特に $\beta \to \infty$ では, ラベルと矛盾のないような予測を与えるパラメータ $\boldsymbol{w}$ の集合の体積である. 以降, 自己平均性により,

$$\mathbb{P}\left[\frac{1}{N}\log V(\mathcal{D}) - \frac{1}{N}\mathbb{E}_{\mathcal{D}}[\log V(\mathcal{D})]\right] \to 0, \quad (N \to \infty)$$

となることを期待し, $\mathbb{E}_{\mathcal{D}}[\log V(\mathcal{D})]$ を評価する. また, 以下の恒等式を用いる.

$$\frac{1}{N}\mathbb{E}_{\mathcal{D}}[\log V(\mathcal{D})] = \frac{1}{N}\lim_{n\to 0}\frac{\partial}{\partial n}\log \mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})] \tag{3}$$

さらに, Gardner 体積を計算するために, $N, M \to \infty$, $M/N \to \alpha$, $\alpha \in (0,\infty)$ の大自由度極限下 (熱力学極限下) で, $\mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})]$ を評価する.

2 レプリカ計算

次の量を具体的に計算していく.

$$\mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})] = \prod_{a=1}^{N}\int d\boldsymbol{w}_a\delta(\|\boldsymbol{w}_a\|^2 - N)\prod_{\mu=1}^{M}\mathbb{E}_{\boldsymbol{x}^\mu,\boldsymbol{w}_0}\left[e^{-\beta\Theta\left(-\frac{\boldsymbol{w}_0^\top \boldsymbol{x}^\mu}{\sqrt{N}}\frac{\boldsymbol{w}_a^\top \boldsymbol{x}^\mu}{\sqrt{N}}\right)}\right] \tag{4}$$

さらに, 特徴量 $\boldsymbol{x}^\mu$ が, 独立なガウス分布であることを考慮すると次の等式が得られる.

$$\prod_{\mu=1}^{M} \mathbb{E}_{\boldsymbol{x}^\mu, \boldsymbol{w}_0}\left[e^{-\beta\Theta\left(\frac{\boldsymbol{w}_0^\top \boldsymbol{x}^\mu}{\sqrt{N}} \frac{\boldsymbol{w}_a^\top \boldsymbol{x}^\mu}{\sqrt{N}}\right)}\right] = \left(\mathbb{E}_{\boldsymbol{x}, \boldsymbol{w}_0}\left[e^{-\beta\Theta\left(-\frac{\boldsymbol{w}_0^\top \boldsymbol{x}^\mu}{\sqrt{N}} \frac{\boldsymbol{w}_a^\top \boldsymbol{x}^\mu}{\sqrt{N}}\right)}\right]\right)^M$$
$$= \left(\int d\boldsymbol{u} \mathcal{N}(\boldsymbol{u}|\boldsymbol{0}_{n+1}, \Sigma_{n+1\times n+1}) e^{-\beta\Theta(u_0 u_a)}\right)^M$$

ここで, $\boldsymbol{u} = (\boldsymbol{w}_a^\top \boldsymbol{x}/\sqrt{N})_{a=0}^n \in \mathbb{R}^{n+1}$ と定義し, $\Sigma_{n+1\times n+1}$ を次のように定義した;

$$\Sigma_{n+1\times n+1} = \begin{pmatrix} \rho & \boldsymbol{m}^\top \\ \boldsymbol{m} & \boldsymbol{Q} \end{pmatrix}, \ \ \rho \equiv \frac{\|\boldsymbol{w}_0\|^2}{N} = 1, \ \ \boldsymbol{m} \equiv \left(\frac{\boldsymbol{w}_0^\top \boldsymbol{w}_a}{N}\right) \in \mathbb{R}^n, \ \ \boldsymbol{Q} \equiv \left(\frac{\boldsymbol{w}_a^\top \boldsymbol{w}_b}{N}\right) \in \mathbb{R}^{n\times n}.$$

この結果を用いると, 次のように整理することができる.

$$\mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})] = \int d\Sigma \left(\int d\boldsymbol{u} \mathcal{N}(\boldsymbol{u}|\boldsymbol{0}_{n+1}, \Sigma_{n+1\times n+1}) e^{-\beta\Theta(u_0 u_a)}\right)^M \prod_a d\boldsymbol{w}_a \prod_{a,b=0} \delta(N\Sigma_{ab} - \boldsymbol{w}_a^\top \boldsymbol{w}_b) \tag{5}$$

ここで, $d\Sigma$ は, $d\Sigma \equiv d\rho \prod_{a=1} dm_a \prod_{a\neq b} dQ_{ab}$ と定義した. 以降, 解析の説明をしやすくするために,

$$\mathcal{T}(\Sigma) = \frac{1}{N}\log\left(\int d\boldsymbol{u} \mathcal{N}(\boldsymbol{u}|\boldsymbol{0}_{n+1}, \Sigma_{n+1\times n+1}) e^{-\beta\Theta(u_0 u_a)}\right)^M$$
$$\mathcal{S}(\Sigma) = \frac{1}{N}\log\int \prod_a d\boldsymbol{w}_a \prod_{a,b=0} \delta(N\Sigma_{ab} - \boldsymbol{w}_a^\top \boldsymbol{w}_b)$$

を導入し, $\mathcal{T}(\Sigma), \mathcal{S}(\Sigma)$ をそれぞれエネルギー項, エントロピー項と呼ぶことにする[*2]. この二つにより形式的に,

$$\mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})] = \int d\Sigma e^{N(\mathcal{T}(\Sigma)+\mathcal{S}(\Sigma))} \tag{6}$$

と表せる. ここからは, エントロピー項, エネルギー項をそれぞれ計算していく.

2.1 エントロピー項の計算

まず, 以下のエントロピー項を評価する.

$$\mathcal{S}(\boldsymbol{\Sigma}) = \frac{1}{N}\ln\int \prod_{a=0}^{n} d\boldsymbol{w}_a \prod_{a,b=0} \delta(N\Sigma_{ab} - \boldsymbol{w}_a^\top \boldsymbol{w}_b). \tag{7}$$

まず log の中身を delta 関数のフーリエ変換表示

$$N\delta(Nq_{ab} - \boldsymbol{w}_a^\top \boldsymbol{w}_b) = \frac{N}{4\pi}\int_{-\infty i}^{+\infty i} d\tilde{q}_{ab} e^{-\frac{\tilde{q}_{ab}}{2}(Nq_{ab} - \boldsymbol{w}_a^\top \boldsymbol{w}_b)} \tag{8}$$
$$N\delta(Nm_a - \boldsymbol{w}_a^\top \boldsymbol{w}_0) = \frac{N}{2\pi}\int_{-\infty i}^{+\infty i} d\tilde{m}_a e^{\tilde{m}_a(Nm_a - \boldsymbol{w}_a \boldsymbol{w}_0)} \tag{9}$$
$$N\delta(N - \|\boldsymbol{w}_a\|^2) = \frac{N}{4\pi}\int_{-\infty i}^{+\infty i} d\tilde{Q}_a e^{\frac{\tilde{Q}_a}{2}(N - \|\boldsymbol{w}_a\|^2)} \tag{10}$$

[*2] この項をエネルギー項, エントロピー項と呼ぶことは慣習であり, 特に深い意味はないはず. ただ, エネルギー項とエントロピー項で解析指針が異なるので別々に解析すると解析しやすいのは確かである.

を用いて, エントロピー項は次のように整理する;

$$
\begin{aligned}
S(\Sigma) \asymp \frac{1}{N}\log\int\prod_{a\neq b}d\tilde{q}_{ab}\prod_{a=1}d\tilde{m}_a\prod_{a=1}d\tilde{Q}_a e^{N\left(-\frac{1}{2}\sum_{a\neq b}\tilde{q}_{ab}q_{ab}+\frac{1}{2}\sum_{a=1}\tilde{Q}_a-\sum_{a=1}\tilde{m}_a m_a\right)} \\
\times\int\prod_{a=1}^{n}d\boldsymbol{w}_a\int\delta(N-\|\boldsymbol{w}_0\|^2)d\boldsymbol{w}_0 e^{\sum_{a\neq b}\frac{\tilde{q}_{ab}}{2}\boldsymbol{w}_a^\top\boldsymbol{w}_b+\sum_{a=1}\tilde{m}_a\boldsymbol{w}_a^\top\boldsymbol{w}_0-\frac{1}{2}\sum_{a=1}\tilde{Q}_a\|\boldsymbol{w}_a\|^2}
\end{aligned}
\tag{11}
$$

ここで, 熱力学極限では, 指数関数の肩が $e^{O(nN)}$ の項だけが効いてくるため, $\asymp$ を, “そのオーダのみ注目すると等しい” という意味で定義した[*3]. ここで, 次のレプリカ対称性を仮定する;

$$
\tilde{q}_{ab}=\tilde{q},\quad \tilde{m}_a=\tilde{m},\quad \tilde{Q}_a=\tilde{Q},\quad q_{ab}=q,\quad m_a=m.
$$

この仮定から次のように表せる;

$$
\begin{aligned}
S(q,m) \asymp \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)} \\
\times\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\prod_{i=1}^{N}\int\prod_{a}dw_{a,i}e^{\frac{\tilde{q}}{2}\sum_{a\neq b}w_{a,i}w_{b,i}+\tilde{m}\sum_{a=1}w_{a,i}w_{0,i}-\frac{1}{2}\sum_{a=1}\tilde{Q}w_{a,i}^2}
\end{aligned}
$$

さらに, $e^{\frac{a^2}{2}}=\int Dze^{az}$ を用いると次のように整理できる.

$$
\begin{aligned}
S(q,m) \asymp \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)+\frac{N}{2}\tilde{Q}_0} \\
\times\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\int Dz\left(\prod_{i=1}^{N}\int dw_i e^{-\frac{1}{2}(\tilde{Q}+\tilde{q})w_i^2+(\sqrt{\tilde{q}}z+\tilde{m}w_{0,i})w_i}\right)^n
\end{aligned}
$$

さらに, 計算を進めると

$$
\begin{aligned}
S(q,m) &\asymp \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)+\log\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\int Dz\left(\prod_{i=1}^{N}\int dw_ie^{-\frac{1}{2}(\tilde{Q}+\tilde{q})w_i^2+(\sqrt{\tilde{q}}z+\tilde{m}w_{0,i})w_i}\right)^n} \\
&= \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)+n\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\int Dz\log\prod_{i=1}^{N}\int dw_ie^{-\frac{1}{2}(\tilde{Q}+\tilde{q})w_i^2+(\sqrt{\tilde{q}}z+\tilde{m}w_{0,i})w_i}} \\
&= \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)+n\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\int Dz\log\left(\frac{2\pi}{\tilde{Q}+\tilde{q}}\right)^{\frac{N}{2}}\prod_{i=1}^{N}e^{\frac{\sqrt{\tilde{q}}z^2+2\sqrt{\tilde{q}}\tilde{m}zw_{0,i}+\tilde{m}^2w_{0,i}^2}{2(\tilde{Q}+\tilde{q})}}} \\
&= \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m\right)+n\int d\boldsymbol{w}_0\delta(N-\|\boldsymbol{w}_0\|^2)\int Dz\log\left(\frac{2\pi}{\tilde{Q}+\tilde{q}}\right)^{\frac{N}{2}}e^{\frac{N\sqrt{\tilde{q}}z^2+2\sqrt{\tilde{q}}\tilde{m}z\sum_i w_{0,i}+\tilde{m}^2\|\boldsymbol{w}_0\|^2}{2(\tilde{Q}+\tilde{q})}}} \\
&= \frac{1}{N}\log\int d\tilde{q}d\tilde{m}d\tilde{Q}e^{Nn\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m+\frac{1}{2}\log\left(\frac{2\pi}{\tilde{Q}+\tilde{q}}\right)+\frac{\sqrt{\tilde{q}}+\tilde{m}^2}{2(\tilde{Q}+\tilde{q})}\right)} \\
&\asymp \mathrm{extr}_{\tilde{Q},\tilde{q},\tilde{m}}\left(\frac{1}{2}(\tilde{q}q+\tilde{Q})-\tilde{m}m+\frac{1}{2}\log\left(\frac{2\pi}{\tilde{Q}+\tilde{q}}\right)+\frac{\tilde{q}+\tilde{m}^2}{2(\tilde{Q}+\tilde{q})}\right)
\end{aligned}
$$

*3 熱力学極限で消えてしまう係数を書くのをサボっているだけで, こういう記号を使うのは良くない気がする. レプリカ解析の論文とかを見るとここを普通に等号で結んでいる論文も多い... ただ, 後でできるだけこの記法を使わないように直す.

鞍点条件から次の関係式が得られる;

$$\partial_{\tilde{q}} : q - \frac{\tilde{q} + \tilde{m}^2}{(\tilde{Q} + \tilde{q})^2} = 0$$
$$\partial_{\tilde{Q}} : 1 - \frac{1}{\tilde{Q} + \tilde{q}} - \frac{\tilde{q} + \tilde{m}^2}{(\tilde{Q} + \tilde{q})^2} = 0$$
$$\partial_{\tilde{m}} : m - \frac{\tilde{m}}{\tilde{Q} + \tilde{q}} = 0$$

この関係式から次の等式が得られる.

$$1 - q = \frac{1}{\tilde{Q} + \tilde{q}}, \quad \tilde{m} = \frac{m}{1-q}, \quad \tilde{q} = \frac{q - m^2}{(1-q)^2}, \quad \tilde{Q} = \frac{1 - 2q + m^2}{(1-q)^2}, \quad \frac{\tilde{q} + \tilde{m}^2}{2(\tilde{Q} + \tilde{q})} = \frac{q}{2(1-q)}$$

この結果から, $\tilde{Q}, \tilde{q}, \tilde{m}$ を削除すると

$$S(q, m) = \frac{1}{2}\log(1-q) + \frac{1 - m^2}{2(1-q)} + \frac{1}{2}\log 2\pi$$

となる.

2.2 Energy 項の計算

次に, 以下の Energy 項を評価する.

$$\mathcal{T}(\Sigma) = \frac{1}{N}\log\left(\int d\boldsymbol{u}\mathcal{N}(\boldsymbol{u}|\boldsymbol{0}_{n+1}, \Sigma_{n+1\times n+1})e^{-\beta\Theta(u_0 u_a)}\right)^M$$

ここで, レプリカ対称性の仮定の元では, 確率変数 $\boldsymbol{u} \in \mathbb{R}^{n+1}$ は以下のように表せることに注目する:

$$u_0 = \sqrt{1 - \frac{m^2}{q}}x_0 + \frac{m}{\sqrt{q}}z, \quad u_a = \sqrt{1-q}x_a + \sqrt{q}z, \quad \forall a = 0, \ldots, n, \ x_a \sim \mathcal{N}(0,1), \ z \sim \mathcal{N}(0,1).$$

実際, 次のように確認することができる:

$$\mathbb{E}_{x_0, z}[u_0 u_0] = 1, \quad \mathbb{E}_{x_0, x_a, z}[u_0 u_a] = m, \quad \mathbb{E}_{x_a, x_b, z}[u_a u_b] = q.$$

この結果をを代入すると次のように計算できる.

$$\begin{aligned}
\mathcal{T}(\Sigma) &\asymp \alpha n \int DzDx_0 \ln \int Dx \exp\left(-\beta\Theta\left[-\left(\sqrt{1 - \frac{m^2}{q}}x_0 + \frac{m}{\sqrt{q}}z\right)\left(\sqrt{1-q}x + \sqrt{z}\right)\right]\right) \\
&= \alpha n \int Dz\left(\int_{-\infty}^{-\frac{m}{\sqrt{q-m^2}}z} Dx_0 \ln\left(\int_{-\sqrt{\frac{q}{1-q}}z}^{\infty} Dxe^{-\beta} + \int_{-\infty}^{-\sqrt{\frac{q}{1-q}}z} Dx\right)\right. \\
&\quad \left. + \int_{-\frac{m}{\sqrt{q-m^2}}z}^{\infty} Dx_0 \ln\left(\int_{-\infty}^{-\sqrt{\frac{q}{1-q}}z} Dxe^{-\beta} + \int_{-\sqrt{\frac{q}{1-q}}z}^{\infty} Dx\right)\right) \\
&= 2\alpha n \int DzH\left(\frac{m}{\sqrt{q-m^2}}z\right)\ln\left[H\left(\sqrt{\frac{q}{1-q}}z\right) + e^{-\beta}\left(1 - H\left(\sqrt{\frac{q}{1-q}}z\right)\right)\right] \\
&= 2\alpha n \int DzH\left(\frac{m}{\sqrt{q-m^2}}z\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right), \quad \beta \to \infty
\end{aligned}$$

2.3 鞍点方程式

これまで計算した entropy 項と energy 項を Eq. (6) に代入すると

$$\frac{1}{N}\log\mathbb{E}_{\mathcal{D}}[V^n(\mathcal{D})] = n\mathrm{extr}_{q,m}\left(2\alpha\int DzH\left(\frac{m}{\sqrt{q-m^2}}z\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right)+\frac{1}{2}\log(1-q)+\frac{1-m^2}{2(1-q)}+\frac{1}{2}\log 2\pi\right)$$

と表せる. さらに, Eq. (3) に代入すると次のように表せる:

$$\frac{1}{N}\mathbb{E}_{\mathcal{D}}[\log V(\mathcal{D})] = \mathrm{extr}_{q,m}\left(2\alpha\int DzH\left(\frac{m}{\sqrt{q-m^2}}z\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right)+\frac{1}{2}\log(1-q)+\frac{1-m^2}{2(1-q)}+\frac{1}{2}\log 2\pi\right)$$

ここで, 推定値 $\boldsymbol{w}$ の分布と teacher の重みの事後分布が等しいことから, $q=m$ となる*4. この事実から, 定数項を除いて

$$\frac{1}{N}\mathbb{E}_{\mathcal{D}}[\log V(\mathcal{D})] \propto \mathrm{extr}_{q}\left(2\alpha\int DzH\left(\sqrt{\frac{q}{1-q}}z\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right)+\frac{1}{2}\log(1-q)+\frac{q}{2}\right) \tag{12}$$

と簡略化できる. 以降, 鞍点条件を計算していく. そのために次の量を導入しておく:

$$\xi \equiv \sqrt{\frac{q}{1-q}}z,\quad \frac{d\xi}{dq} = \frac{z}{2(1-q)^2\sqrt{\frac{q}{1-q}}}$$

この量を用いて, Eq. (12) の右辺を q に関して微分して 0 の条件は次のように表せる:

$$2\alpha\frac{d}{dq}\int DzH(\xi)\ln H(\xi) = \frac{q}{2(1-q)} \tag{13}$$

左辺は次のように計算できる;

$$\begin{aligned}
2\alpha\frac{d}{dq}\int DzH(\xi)\ln H(\xi) &= 2\alpha\int DzH'(\xi)\frac{d\xi}{dq}\ln H(\xi)+\int DzH(\xi)\frac{1}{H(\xi)}\frac{d\xi}{dq}\\
&= -2\alpha\int Dz\frac{1}{\sqrt{2\pi}}e^{-\frac{q}{1-q}z^2}\left(\frac{z}{2(1-q)^2\sqrt{\frac{q}{1-q}}}\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right)\\
&= -\frac{2\alpha}{\sqrt{2\pi}}\int dz\frac{1}{\sqrt{2\pi}}e^{-\left(\frac{1}{1-q}\right)z^2}\left(\frac{z}{2(1-q)^2\sqrt{\frac{q}{1-q}}}\right)\ln H\left(\sqrt{\frac{q}{1-q}}z\right)\\
&= -\frac{2\alpha}{\sqrt{2\pi}}\frac{\sqrt{1-q}}{2(1-q)\sqrt{q}}\int Dtt\ln H\left(\sqrt{q}t\right),\quad \frac{z}{\sqrt{1-q}}=t\\
&= \frac{\alpha}{\pi}\frac{\sqrt{1-q}}{2(1-q)}\int Dt\frac{e^{-\frac{1}{2}qt^2}}{H\left(\sqrt{q}t\right)}.
\end{aligned}$$

まとめると鞍点条件は次のように表せる:

$$q = \frac{\alpha}{\pi}\sqrt{1-q}\int Dt\frac{e^{-\frac{1}{2}qt^2}}{H\left(\sqrt{q}t\right)} = \frac{\alpha}{\pi}\sqrt{\frac{1-q}{q}}\int D\gamma\frac{e^{-\frac{1}{2q}\gamma^2}}{H(\gamma)},\quad \gamma=\sqrt{q}t \tag{14}$$

この方程式をとき, 次の汎化誤差

$$\epsilon_g = \frac{1}{\pi}\cos^{-1}(q)$$

を評価した結果を Fig. 1に示す.

---

*4 この結果は, 鞍点方程式を解くと形式的に求まる. 今回は, 楽をするために最初にこの結果を代入して, q に関する鞍点条件を計算する.

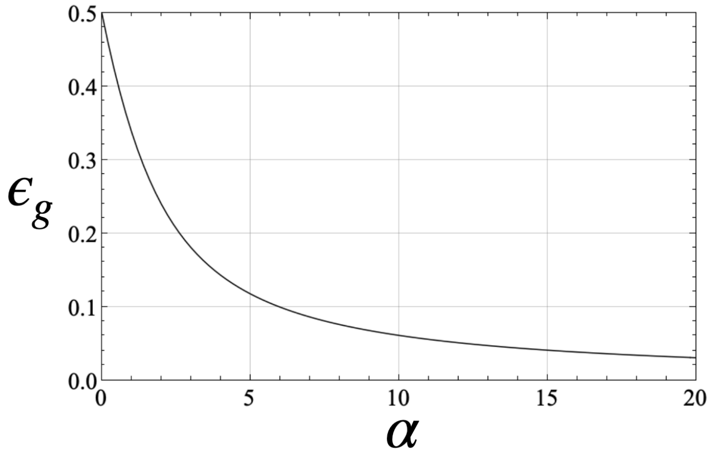

図 1 汎化誤差 ϵ_g の α 依存性.

2.4 漸近挙動

学習データ数が大きいときの極限を考えるために, $q = 1 - \epsilon$ とおき, Eq. (14) を展開する.

$$1 - \epsilon = \frac{\alpha}{\pi}\left(\sqrt{\epsilon} + O(\epsilon^{\frac{3}{2}})\right) \int D\gamma \frac{1}{H(\gamma)} \left(e^{-\frac{1}{2}\gamma^2} - \frac{1}{2}e^{-\frac{1}{2}\gamma^2}\gamma^2\epsilon + O(\epsilon^2)\right)$$

$\sqrt{\epsilon}$ までの展開に着目すると次の関係式が得られる:

$$\epsilon \sim \frac{\pi^2}{\alpha^2 c^2}, \quad c = \int D\gamma \frac{e^{-\frac{1}{2}\gamma^2}}{H(\gamma)}$$

ゆえに, 汎化誤差の漸近的な振る舞いは次のように計算できる;

$$\epsilon_g \sim \frac{1}{\pi}\sqrt{2\epsilon} \sim \frac{\sqrt{2}}{\alpha c} = \frac{0.625}{\alpha}$$

この結果から, 学習データ数が十分大きいとき, 汎化誤差は, 学習データ数の逆数に比例して減少することが明らかとなった.

参考文献

Andreas Engel and Christian Van den Broeck. Statistical mechanics of learning. Cambridge University Press, 2001.
秀稔 西森. 新物理学選書スピングラス理論と情報統計力学. 岩波書店, 2 2016. ISBN 978-4007303654.