§4. 多値分類

$(X, Y) \sim \mathcal{D} \in \Delta(\mathcal{X} \times \mathcal{Y})$　特に $\mathcal{Y} = \{1, \cdots, K\}$

目標：誤分類率 $\Pr(Y \neq h(X))$ を最小にする分類器 $h : \mathcal{X} \to \mathcal{Y}$ の学習.

各クラス $y \in \mathcal{Y}$ に対して，予測器 $f_y : \mathcal{X} \to \mathbb{R}$ を学習し，$h(x) = \underset{y \in \mathcal{Y}}{\operatorname{argmax}}\, f_y(x)$ で予測.

（他にも one-vs-one, ECOC などの定式化あり）

（定義上は集合だが，ここでは適当に & 決定的に tie-breaking する）

$$R[h] := \Pr(Y \neq h(X)) = \mathop{\mathbb{E}}_{(X,Y)\sim\mathcal{D}}\left[\ell(Y, h(X))\right] \quad \text{where 0-1損失 } \ell(y, \hat{y}) := \mathbb{1}[y \neq \hat{y}]$$

$$= \mathop{\mathbb{E}}_{X}\left[\left\langle \eta(X), \ell_{h(X)} \right\rangle\right]$$

条件つき分布　　条件つき 0-1 リスク.

$$\eta(x) = \begin{bmatrix} \Pr(Y=1 \mid X=x) \\ \vdots \\ \Pr(Y=k \mid X=x) \end{bmatrix} \in \Delta_K \qquad \ell_{\hat{y}} := \begin{bmatrix} \ell(1, \hat{y}) \\ \vdots \\ \ell(k, \hat{y}) \end{bmatrix} \in \mathbb{R}^K$$

定義. Bayes リスク $R^* := \inf_{h \in \mathcal{M}(\mathcal{X}, \mathcal{Y})} R(h) = \mathop{\mathbb{E}}_{X}\left[\underbrace{\min_{\hat{y} \in \mathcal{Y}} \left\langle \eta(X), \ell_{\hat{y}} \right\rangle}_{\text{条件つき Bayes リスク}}\right]$

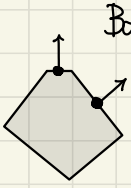
- 多値の場合の代理損失 $\phi : \mathcal{Y} \times \mathbb{R}^K \to \mathbb{R}$

定義 ϕ-リスク $R_\phi(f) := \mathbb{E}\left[\phi(Y, f(X))\right] = \mathop{\mathbb{E}}_{X}\left[\overbrace{\left\langle \eta(X), \phi(f(X)) \right\rangle}^{\text{条件つき } \phi\text{-リスク}}\right]$

$\eta \in \Delta_K$　　$\phi(\alpha) = \begin{bmatrix} \phi(1, \alpha) \\ \vdots \\ \phi(K, \alpha) \end{bmatrix} \in \mathbb{R}^K$

Bayes ϕ-リスク $R_\phi^* := \inf_{f \in \mathcal{M}(\mathcal{X}, \mathbb{R})} R_\phi(f) = \mathop{\mathbb{E}}_{X}\left[\overbrace{\inf_{\alpha \in \mathbb{R}^K} \left\langle \eta(X), \phi(\alpha) \right\rangle}^{\text{条件つき Bayes } \phi\text{-リスク}}\right]$

$= \inf_{z \in S_\phi} \left\langle \eta(X), z \right\rangle$: 凸集合 S_ϕ の支持関数になっている



$S_\phi = \operatorname{conv}(\phi(\mathbb{R}^K))$

- 代理損失の例.

① Crammer-Singer 損失. $\phi(y,\alpha) := \left[1-\left(\alpha_y - \max_{k\neq y}\alpha_k\right)\right]_+$ where $[\cdot]_+ := \max\{0,\cdot\}$

二値分類のヒンジ損失 $\phi(\alpha)=[1-\alpha]_+$ の拡張. ↳ 気持ち: $\forall k\neq y$ について $\alpha_y > \alpha_k$ が成立してほしい.

② One-vs-Rest (OvR) 損失 $\phi(y,\alpha) := \bar{\phi}(\alpha_y) + \sum_{k\neq y}\bar{\phi}(-\alpha_k)$ where $\bar{\phi}:\mathbb{R}\to\mathbb{R}$. ($\bar{\phi}$: 二値用の損失)

y を正例, それ以外 k を負例 とみなす.

- Q. 代理損失最小化は妥当?

Def. ϕ が ℓ-判別適合的 $\overset{\text{def.}}{\iff}$ $\forall \eta\in\triangle_k$ に対し次が成立.

$$\inf_{\alpha\in\mathbb{R}^k}\left\{\langle \eta,\phi(\alpha)\rangle : \underbrace{\operatorname*{argmax}_{\hat{y}\in Y}\alpha_{\hat{y}} \notin \operatorname*{argmin}_{\hat{y}\in Y}\langle \eta,\ell_{\hat{y}}\rangle}_{\text{予測 } \alpha \text{ が suboptimal.}}\right\} > \underbrace{\inf_{\alpha\in\mathbb{R}^k}\langle \eta,\phi(\alpha)\rangle}_{\text{Bayes } \phi\text{-リスク}}$$

⇝ 二値分類における $\underbrace{C^-_{\phi,\eta}}_{\text{Subopt なリスク}} > \underbrace{C^*_{\phi,\eta}}_{\text{Bayes リスク}}$ の拡張.

- Crammer-Singer は ℓ-判別適合的ではない.

CS 損失を一般化して $\phi(y,\alpha) := \bar{\phi}\left(\alpha_y - \max_{k\neq y}\alpha_k\right)$ を考える. (多値マージン. vs. 二値マージンな α)

Prop. 4.1. $\bar{\phi}$ は 凸かつ非増加を仮定. このとき CS 損失は ℓ-判別適合的でない.

Proof. $\eta \in \Delta_K$ s.t. $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_K$ としても WLOG. $C_{\phi,\eta}(\alpha) := \langle \eta, \phi(\alpha) \rangle$.

$\inf_{\alpha \in \mathbb{R}^K} \langle \eta, \phi(\alpha) \rangle$ は $\mathcal{A} := \{ \alpha \in \mathbb{R}^K : \alpha_1 \geq \alpha_2 = \cdots = \alpha_K = 0 \}$ に制限してもよい.

まず α として $\alpha_1 \geq \alpha_k$ $(\forall k \neq 1)$ をとる. (∃k≠1 s.t. $\alpha_1 < \alpha_k$ だとしたら判別適合性の定義に矛盾するので考えない)

このとき,

$$C_{\phi,\eta}(\alpha) = \eta_1 \bar{\phi}\Big(\underbrace{\alpha_1 - \max_{k \neq 1} \alpha_k}_{\leq \alpha_1 - \alpha'}\Big) + \underbrace{\sum_{y=2}^{K} \eta_y \bar{\phi}(\alpha_y - \alpha_1)}_{= (1-\eta_1) \sum_{y=2}^{K} \frac{\eta_y}{1-\eta_1} \bar{\phi}(\alpha_y - \alpha_1)}$$

$$\geq (1-\eta_1) \bar{\phi}\Big(\underbrace{\sum_{y=2}^{K} \frac{\eta_y}{1-\eta_1} \alpha_y}_{=: \alpha' (\in \mathbb{R})} - \alpha_1 \Big) \quad \text{(Jensen)}$$

$$\geq \eta_1 \bar{\phi}(\alpha_1 - \alpha') + (1-\eta_1) \bar{\phi}(\alpha' - \alpha_1)$$

つまり, $(\alpha_2, \cdots, \alpha_K) \rightsquigarrow (\alpha', \cdots, \alpha')$ としても $C_{\phi,\eta}$ の最小化 in $\alpha \in \mathbb{R}^K$ に影響しない.

また, $(\alpha_1, \cdots, \alpha_K)$ は定数を足しても影響しないので, $\alpha' = 0$ としてよい. □

ここで η として $\eta_1 < \frac{1}{2}$ なるものを考える. (K≥3(多値)なので可能.) $\alpha \in \mathcal{A}$ に対して,

$$C_{\phi,\eta}(\alpha) = \eta_1 \bar{\phi}(\alpha_1) + (1-\eta_1) \bar{\phi}(-\alpha_1) \quad \text{(K-dimから1-dimの帰着)}$$

$$= \eta_1 \underbrace{\big(\bar{\phi}(\alpha_1) + \bar{\phi}(-\alpha_1)\big)}_{\geq 2\bar{\phi}(0) \text{ (Jensen)}} + \underbrace{(1-2\eta_1)}_{>0} \underbrace{\bar{\phi}(-\alpha_1)}_{\geq \bar{\phi}(0) \ (\because \bar{\phi} \text{ 非増加})}$$

$$\geq \bar{\phi}(0).$$

$$= C_{\phi,\eta}(0).$$

したがって $\inf_{\alpha \in \mathbb{R}^K} C_{\phi,\eta}(\alpha) = \inf_{\alpha \in \mathcal{A}} C_{\phi,\eta}(\alpha)$ の最小化は $\alpha_1^* = \alpha_2^* = \cdots = \alpha_K^*$ で達成.

与 $\operatorname{argmax}_{y \in Y} \alpha_y$ の tie-breaking は任意なので, $\operatorname{argmax}_{y \in Y} g^*(x) \notin \operatorname{argmin}_{y \in Y} \langle \eta, \ell_y \rangle$ があり得る.

$\Rightarrow$ ϕ は判別適合的でない. □

- Remark : $\eta_1 \geq \frac{1}{2}$ (i.e. $\max_{y \neq y} \eta_y \geq \frac{1}{2}$) を $\mathcal{D}$ (data) に仮定できるなら, CS損失 は 判別適合的.
  "majority assumption"

- OvR $\phi(y,\alpha) = \bar{\phi}(\alpha_y) + \sum_{k \neq y} \bar{\phi}(-\alpha_k)$ は判別適合的.

Prop 4.2. $\bar{\phi}$ は 凸, 下に有界, 可微分, かつ $\bar{\phi}(\alpha) < \bar{\phi}(-\alpha)$ $(\forall \alpha > 0)$ を仮定.
このとき OvR 損失 ϕ は ℓ- 判別適合的.

以下の lemma が重要.

Lem 4.3. 同じ仮定の下で, $\eta_i < \eta_j \Rightarrow \alpha_i^* < \alpha_j^*$ が成立.
ただし, $\alpha^* \in \mathbb{R}^k$ は $C_{\phi,\eta}(\alpha) = \langle \eta, \phi(\alpha) \rangle$ の minimizer.
(順序保存の性質)

Proof of Prop 4.2. WLOG で $\eta_1 < \eta_2$ を仮定して次を示す.

$$\inf_{\alpha \in \mathbb{R}^k} \left\{ \langle \eta, \phi(\alpha) \rangle : \underbrace{\alpha_1 = \max_k \alpha_k}_{\text{suboptimal } \alpha} \right\} > \inf_{\alpha \in \mathbb{R}^k} \langle \eta, \phi(\alpha) \rangle. \quad \text{---} \circledast$$

列 $(\alpha^{(m)})_{m \in \mathbb{N}} (\subseteq \mathbb{R}^k)$ s.t. $\langle \eta, \phi(\alpha^{(m)}) \rangle \xrightarrow{m \to \infty} \inf \left\{ \langle \eta, \phi(\alpha) \rangle : \alpha_1 = \max_k \alpha_k \right\}$ なるものを考える.

$\alpha^{(m)} \to \alpha'$ なる α' を書くと, $\langle \eta, \phi(\alpha') \rangle = \lim_{m \to \infty} \langle \eta, \phi(\alpha^{(m)}) \rangle$.

かつ $\alpha_1' = \max_k \alpha_k'$

Lem 4.3 より $C_{\phi,\eta}$ の minimizer は $\alpha_1^* < \alpha_2^*$ を満たすので, $\circledast$ が成立. □