

Part I: Classification

§1. 二値分類

$(X, Y) \sim \mathcal{D} \in \Delta(X \times Y)$ where $X \in \mathcal{X} (\subseteq \mathbb{R}^d)$ & $Y \in \mathcal{Y} = \{+1, -1\}$.
特徴量 ラベル. positive negative

目標: 分類器 $f: \mathcal{X} \rightarrow \mathbb{R}$ の学習

via 誤分類率 $\Pr(\text{sign}(f(x)) \neq Y)$ の最小化.

$$\text{sign}(\alpha) := \begin{cases} +1 & : \alpha > 0 \\ -1 & : \alpha \leq 0 \end{cases}$$

• 誤分類率

$$\Pr(\text{sign}(f(x)) \neq Y) = \mathbb{E}_{X, Y} [\underbrace{L_{01}(Y, f(x))}_{\substack{\text{0-1損失} \\ \text{ラベル} \nearrow \text{スコア}}}] =: R(f)$$

$$L_{01}(y, \alpha) := \begin{cases} +1 & : \text{sign}(\alpha) \neq y \\ 0 & : \text{sign}(\alpha) = y \end{cases}$$

• $R(f)$ の直接最小化は困難.

例) $f \in \mathcal{F}$ として $\mathcal{F}$ が DNF のとき. NP-hard
 $\mathcal{X}^{\mathbb{R}}$ 仮説空間



Intuition: $R(f) = \mathbb{E} \left[\mathbb{1}_{[-\infty, 0]}(Y f(x)) \right] = \text{Weighted Sum} \left[\underbrace{\begin{matrix} L_{01} \\ 1 \\ 0 \end{matrix}}_{\substack{\text{ラベル} \\ \text{スコア}}} \right]$

f の勾配が a.s. で消失 (泣)

• 凸緩和が必要.

$$R_{\phi}(f) := \mathbb{E} [\phi(Y f(x))] \text{ with } \phi: \mathbb{R} \rightarrow \mathbb{R}.$$

$$\phi: \text{convex} \rightsquigarrow R_{\phi}: \text{convex in } f.$$



- 推定, 近似, 最適化.

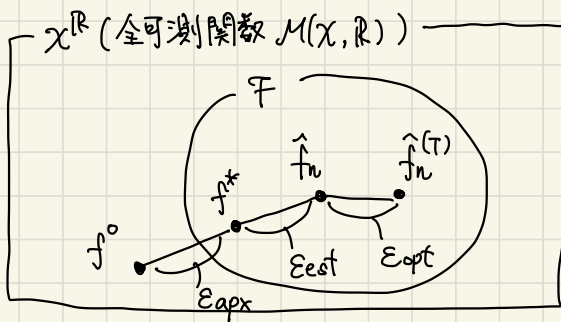
(観測データ)

$R_\Phi(f) = \mathbb{E}[\phi(Yf(x))]$ は finite sample $S = \{(x_i, y_i)\}_{i=1}^n$ による R_Φ は期待リスク. $\mathbb{E}$ を用いて推定する必要あり.

{expected, population}-risk

$\leadsto \hat{R}_\Phi(f) := \frac{1}{n} \sum_{i=1}^n \phi(y_i f(x_i))$ 経験リスク.
empirical risk.

$\leadsto \min_{f \in \mathcal{F}} \hat{R}_\Phi(f)$ を解く: 経験リスク最小化
emp. risk min. = ERM.
仮説空間



$$\hat{f}_n^{(\tau)} = \text{Arg}(\hat{R}_\Phi)$$

$$\hat{f}_n \in \arg\min_{f \in \mathcal{F}} \hat{R}_\Phi(f).$$

$$f^* \in \arg\min_{f \in \mathcal{F}} R_\Phi(f).$$

$$f^o \in \arg\min_{f \in M(x, R)} R_\Phi(f)$$

$\left(\begin{array}{l} E_{opt} \text{ 最適化誤差} \\ E_{est} \text{ 推定誤差} \\ E_{apx} \text{ 近似誤差} \end{array} \right) \begin{array}{l} \mathcal{F} \text{ が小さい} \\ \rightarrow \text{better} \\ \mathcal{F} \text{ が大きい} \\ \rightarrow \text{better} \end{array}$

trade-off.

$\rightarrow$ NOT in focus.

- 損失関数の gap.

解く対象: $\min_{f \in M(x, R)} R_\Phi(f)$

$\left(E_{opt} + E_{est} + E_{apx} \text{ が小さく} \right)$
抑えたい想定

実際に解きたい
対象

$\min_{f \in M(x, R)} R(f)$

gap.

- 例1: 二乗損失 $\phi(\alpha) = (1-\alpha)^2$

最小二乗法: $\min_{w \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \left(\underbrace{y_i}_{\in \{-1, 1\}} - w^T x_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (1 - y_i w^T x_i)^2$

- 例2: ロジスティック損失 $\phi(\alpha) = \log(1 + e^{-\alpha})$.

一般化線形モデル
GLM $Y = 1 \mid X = x \sim \text{Bernoulli}(\mu)$

where $\mu = \underbrace{\psi^{-1}(w^T x)}_{[0, 1]}$; $\psi^{-1}(\alpha) := \frac{1}{1 + e^{-\alpha}}$
逆リンク関数.
(逆ロジットリンク)

→ MLE $\max_{w \in \mathbb{R}^d} \prod_{i=1}^n \underbrace{\left[\psi'(w^T x_i) \right]^{\frac{1+y_i}{2}}}_{y_i=1 \text{ の場合}} \underbrace{\left[1 - \psi'(w^T x_i) \right]^{\frac{1-y_i}{2}}}_{y_i=-1 \text{ の場合}}$
 $\frac{1}{1 + e^{-w^T x_i}} \rightsquigarrow 1 - \frac{1}{1 + e^{-w^T x_i}} = \frac{1}{1 + e^{+w^T x_i}}$

$-\log \rightarrow \min_{w \in \mathbb{R}^d} \sum_{i=1}^n \log(1 + e^{-y_i w^T x_i})$

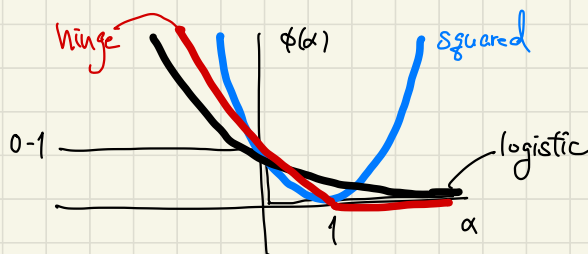
- 例3: ヒンジ損失 $\phi(\alpha) = \max\{0, 1-\alpha\}$

マージン最大化問題 (SVM) 由来.

- 例4: 指数損失 $\phi(\alpha) = e^{-\alpha}$.

ブースティング (特に AdaBoost) 由来.

- その他: Huber loss, ramp loss, ...



§2. 半別適合的損失

Q. $\min_f \underbrace{R_\phi(f)}_{\text{surrogate loss}} \approx \min_f \underbrace{R(f)}_{\text{target loss}}$ の gap は正当化可能?

• Def. $R_\phi(f) = \mathbb{E}[\phi(Yf(x))]$ $R(f) = \mathbb{E} \mathbb{1}[Y \neq \text{sign}(f(x))]$.

Bayes 1/27 $R^* := \inf_{f \in \mathcal{M}(X, \mathbb{R})} R(f)$

Bayes ϕ -1/27 $R_\phi^* := \inf_{f \in \mathcal{M}(X, \mathbb{R})} R_\phi(f)$. R_ϕ の最小化値

• Q': $\forall \mathcal{D} \in \Delta(X, Y) \quad \forall (f_m)_{m \in \mathbb{N}} \subseteq \mathcal{M}(X, \mathbb{R})$,

$$R_\phi(f_m) \rightarrow R_\phi^* \implies R(f_m) \rightarrow R^* \quad ?$$

• Q'': $\forall \mathcal{D}, \forall (f_m)_{m \in \mathbb{N}}$

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \underbrace{R_\phi(f_m) - R_\phi^*}_{\text{surrogate diff.}} < \delta \implies \underbrace{R(f_m) - R^*}_{\text{target diff.}} < \varepsilon$$

Q'' 存在する?

$$J(\varepsilon) := \inf_{\mathcal{D}} \inf_f \left\{ R_\phi(f) - R_\phi^* \mid R(f) - R^* \geq \varepsilon \right\}$$