(§2. 判別適合的損失)

ϕが"判別適合的": $R_\phi(f) \to R_\phi^* \implies R(f) \to R^*$

Thm 2.3 ϕ: convex

ϕが"判別適合的" $\iff$ ϕが"$\alpha=0$で可微分 & $\phi'(0)<0$

- **Proof** ($\Rightarrow$)

ϕの凸性 (at $\alpha=0$) より、$\exists g_1, g_2$ s.t. $\forall \alpha \in \mathbb{R}$, ($\alpha=0$における$\phi$の劣勾配.)

$$\begin{cases} \phi(\alpha) \geq \phi(0) + g_1 \alpha \\ \phi(\alpha) \geq \phi(0) + g_2 \alpha \end{cases}$$

が"成立.

劣勾配不等式 $\phi(\alpha') \geq \phi(\alpha) + \phi'(\alpha)(\alpha' - \alpha)$

背理法: ϕは$\alpha=0$で可微分でない と仮定. $\iff$ $g_1 \neq g_2$

$\eta > \frac{1}{2}$, $\alpha > 0$ を考えると,

$$\begin{aligned} C_{\phi,\eta}(\alpha) &= \eta\phi(\alpha) + (1-\eta)\phi(-\alpha) \\ &\geq \eta(\phi(0) + g_1\alpha) + (1-\eta)(\phi(0) - g_2\alpha) \\ &= \underbrace{((g_1+g_2)\eta - g_2)}_{>0 \text{ by } \eta > \frac{1}{2} \,\&\, g_1 > g_2}\alpha + \phi(0). \\ &> \phi(0) \\ &> C_{\phi,\eta}^* \quad (\text{Lem 2.4}) \end{aligned}$$

矛盾. $\eta \leq \frac{1}{2}$ & $\alpha \leq 0$ のときも同様.

$\Rightarrow$ ϕ の $\alpha=0$ での可微分性 ok.

$\phi'(0) < 0$ を確認する:

$$C_{\phi,\eta}(\alpha) \geq (2\eta - 1)\phi'(0)\alpha + \phi(0)$$

仮に $\eta > \frac{1}{2}$, $\alpha > 0$, $\phi'(0) \geq 0$ とすると $C_{\phi,\eta}(\alpha) > C_{\phi,\eta}^*$ となり同様に矛盾.

$\Rightarrow \phi'(0) < 0$. □

• Proof ($\Leftarrow$)　$\phi'(0)<0$ が仮定.

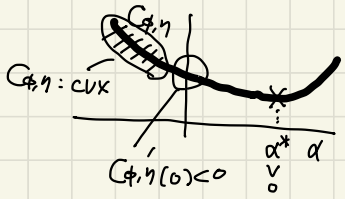
$C_{\phi,\eta}(\alpha) = \eta\phi(\alpha) + (1-\eta)\phi(-\alpha)$.

$C'_{\phi,\eta}(0) = \eta\phi'(0) - (1-\eta)\phi'(0) = (2\eta-1)\phi'(0)$.

$\eta > \frac{1}{2}$ の場合: $C'_{\phi,\eta}(0) < 0 \Rightarrow C_{\phi,\eta}(\alpha)$ は $\alpha = \alpha^* > 0$ で最小化される.

$\eta < \frac{1}{2}$ の場合も同様.

$\Rightarrow C^*_{\phi,\eta} < C^-_{\phi,\eta} \Leftrightarrow$ 判別適合的　□



• まとめ.

α^* ← $C_{\phi,\eta}(\alpha)$ の最小元.

ϕ		$\phi'(0)$	α^*
ヒンジ	$\max\{0, 1-\alpha\}$	-1	$\mathrm{sign}(2\eta-1)$
ロジスティック	$\log(1+e^{-\alpha})$	$-\frac{1}{2}$	$\log\left(\frac{\eta}{1-\eta}\right)$
二乗	$(1-\alpha)^2$	-2	$2\eta-1$

§3. Proper scoring rules

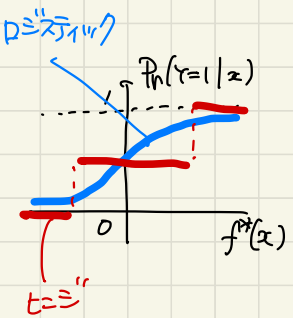
学習の枠組み: ERM $\min_{f\in\mathcal{F}} R_\phi(f)$ ⇝ 分類 $x \mapsto \begin{cases} 1 & : f(x) > 0 \\ -1 & : f(x) \le 0 \end{cases}$

Q. 分類 $f(x)$ の確信度は?

ϕ: ロジスティックの場合. $f^*(x) = \underbrace{\log\left(\frac{\Pr(Y=1|X=x)}{\Pr(Y=-1|X=x)}\right)}_{\text{ロジット}}$

近似 ⇝ $\Pr(Y=1|X=x) = \dfrac{1}{1+e^{-f^*(x)}}$

良さそう?



ϕ: ヒンジの場合. $f^*(x) = \mathrm{sign}(2\Pr(Y=1|X=x)-1)$

⇝ $\Pr(Y=1|X=x) = \begin{cases} 0 & : f^*(x) < -1 \\ \frac{1}{2} & : -1 \le f^*(x) \le 1 \\ 1 & : f^*(x) > 1 \end{cases}$

Def Scoring rule $s: \{\pm 1\} \times [0,1] \to \mathbb{R}$ が proper

def ⇔ $\forall \eta, \hat{\eta} \in [0,1] \quad L(\eta, \eta) \leq L(\eta, \hat{\eta})$

ただし $L(\eta, \hat{\eta}) := \mathop{\mathbb{E}}_{Y \sim \eta} s(Y, \hat{\eta})$ (条件つきリスク)

さらに $L(\eta, \eta) \lneqq L(\eta, \hat{\eta})$ (for $\eta \neq \hat{\eta}$) のとき strictly proper.

η: 正しい条件つき確率. Y: ラベル. $\hat{\eta}$: 条件つき確率 η の推定量

⇝ $\hat{\eta} = \eta$ (推定=真) のときに最適解になってほしい.

- **Proper scoring rule + 逆リンク関数 → Surrogate loss**

逆リンク関数 $\psi^{-1}: \mathbb{R} \to [0,1]$ をひとつとる.

$\alpha \mapsto \hat{\eta}$

このとき $s(y, \psi^{-1}(\alpha)) = \phi(y\alpha)$ と見なせる.

例: log score $\begin{cases} s(1, \hat{\eta}) = -\log \hat{\eta} \\ s(-1, \hat{\eta}) = -\log(1-\hat{\eta}) \end{cases}$

$L(\eta, \hat{\eta}) = \underbrace{-\eta \log \hat{\eta} - (1-\eta)\log(1-\hat{\eta})}_{\eta \text{と} \hat{\eta} \text{の交差エントロピー}}$ は $L(\eta, \hat{\eta}) \geq L(\eta, \eta)$ を満たす. (Gibbs の不等式)

logit link $\psi(\hat{\eta}) = \log\left(\frac{\hat{\eta}}{1-\hat{\eta}}\right) \rightsquigarrow \psi^{-1}(\alpha) = \frac{1}{1+e^{-\alpha}}$

↳ logistic loss $\phi(\alpha) = \log(1+e^{-\alpha}) = s(1, \psi^{-1}(\alpha))$

- sが "properである条件"

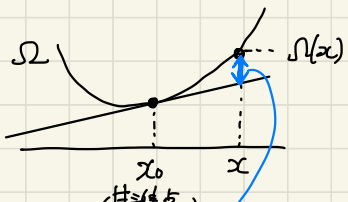
A. Bregman ダイバージェンス で書けること

Def. 凸関数 $\Omega: \mathbb{R} \to \overline{\mathbb{R}}$ に対して、以下を Bregman ダイバージェンスという：

$$B_\Omega(x \| x_0) := \Omega(x) - \Omega(x_0) - g(x_0)\cdot(x - x_0)$$

ただし $g(x_0) \in \partial\Omega(x_0)$ （各点 $x_0 \in \mathrm{dom}\,\Omega$ において）

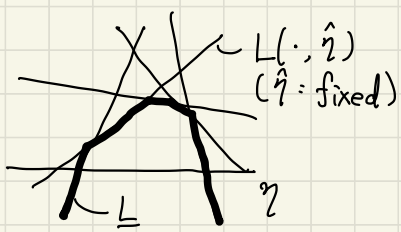
$\partial\Omega$：Ω の劣勾配.



↝ B_Ω は x_0 での Taylor 1次展開の誤差を x で評価.

Def. 条件つきリスク $L(\eta, \hat\eta) := \eta\, s(1, \hat\eta) + (1-\eta)\, s(-1, \hat\eta)$

条件つき Bayes リスク $\underline{L}(\eta) := \inf_{\hat\eta \in [0,1]} L(\eta, \hat\eta)$.



Note: $\underline{L}$ は常に凹関数 ← 参照

Leonald J. Savage (1971)

Thm. 3.1 (Savage) sが正則を仮定、つまり

$s(1, \hat\eta) < \infty$ $(\forall \hat\eta \neq 0)$ かつ $s(-1, \hat\eta) < \infty$ $(\forall \hat\eta \neq 1)$

このとき、sは (strictly) properであることと以下は同値：

ある (strictly) convex Ω が存在し、

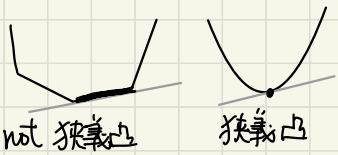
$\forall \hat\eta \in [0,1]$, $\exists\, \hat g \in \partial\Omega(\hat\eta)$ s.t. $L(\eta, \hat\eta) = -\Omega(\hat\eta) - \hat g \cdot (\eta - \hat\eta)$ $(\forall \eta)$

- $\underline{L(\eta, \hat\eta) - \underline{L}(\eta)}$ $\overset{\Omega = -\underline{L}}{=}$ $\underbrace{(-\underline{L})(\eta) - (-\underline{L})(\hat\eta) - \hat g \cdot (\eta - \hat\eta)}_{B_{-\underline{L}}(\eta \| \hat\eta)}$

regret = 真の η に対する $\hat\eta$ の suboptimality

- strictly proper ⇔ 狭義凸

↝ 凸性を使って properness が検証可能.



- **Proof ($\Rightarrow$)** S : proper は仮定.

Ω として $\Omega(\eta) = -L(\eta,\eta) = -\underline{L}(\eta)$ $\left[= -\eta s(1,\eta) - (1-\eta) s(-1,\eta)\right)$ を選ぶ

このとき $-\Omega(\eta) \le L(\eta,\hat{\eta})$ (s: proper)

$= \eta s(1,\hat{\eta}) + (1-\eta) s(-1,\hat{\eta})$

$= \underbrace{\hat{\eta} s(1,\hat{\eta}) + (1-\hat{\eta}) s(-1,\hat{\eta})}_{-\Omega(\hat{\eta})} + (\eta-\hat{\eta})\left(s(1,\hat{\eta}) - s(-1,\hat{\eta})\right)$

$\rightsquigarrow \Omega(\eta) \ge \Omega(\hat{\eta}) + \left(s(-1,\hat{\eta}) - s(1,\hat{\eta})\right)\cdot(\eta-\hat{\eta})$

$\rightsquigarrow$ 劣勾配の定義より $s(-1,\hat{\eta}) - s(1,\hat{\eta}) \in \partial\Omega(\hat{\eta})$.

次に $\hat{g} = s(-1,\hat{\eta}) - s(1,\hat{\eta})$ を所望の式の RHS に代入:

$-\Omega(\hat{\eta}) - \left(s(-1,\hat{\eta}) - s(1,\hat{\eta})\right)(\eta-\hat{\eta})$

$= \left[\hat{\eta} s(1,\hat{\eta}) + (1-\hat{\eta}) s(-1,\hat{\eta})\right] - (\eta-\hat{\eta}) s(-1,\hat{\eta}) + (\eta-\hat{\eta}) s(1,\hat{\eta})$.

$= \eta s(1,\hat{\eta}) + (1-\eta) s(-1,\hat{\eta})$

$= L(\eta,\hat{\eta})$. □

- **Proof ($\Leftarrow$)** 凸関数 Ω が以下を満たすのが仮定:

$\forall \hat{\eta} \in [0,1]$. $\exists \hat{g} \in \partial\Omega(\hat{\eta})$ s.t. $L(\eta,\hat{\eta}) = -\Omega(\hat{\eta}) + \hat{g}\cdot(\eta-\hat{\eta})$ $(\forall \eta)$

このとき, $L(\eta,\eta) = -\Omega(\eta)$

$\le -\Omega(\hat{\eta}) - \hat{g}\cdot(\eta-\hat{\eta})$ (Ω: convex)

$= L(\eta,\hat{\eta})$. □