(§3. Proper scoring rules)

scoring rule

$L(\eta, \hat{\eta}) := \underset{Y \sim \eta}{\mathbb{E}}\, s(Y, \hat{\eta})$ — $s(Y, \hat{\eta})$: クラス確率の推定 $\in [0,1]$, 観測ラベル $\in \{\pm 1\}$, 真 $\in [0,1]$

が"proper" $\overset{\text{def}}{\Leftrightarrow}$ $L(\eta, \hat{\eta}) \geq L(\eta, \eta)$ $\forall \eta, \hat{\eta}$

Thm 3.1 (Savage) s が正則を仮定.

s が"(strictly) proper

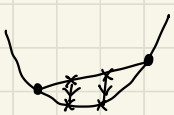
$\Leftrightarrow$ Ω: (strictly) convex が存在して,

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

• **Proof (strictness)**

[$\Rightarrow$] s: strictly proper を仮定して, $\Omega(\eta) = -L(\eta, \eta)$ が狭義凸を示す.

任意に $t \in (0,1)$ をとり,



$$
\begin{aligned}
-\Omega(t\eta + (1-t)\eta') &= L(t\eta + (1-t)\eta',\ t\eta + (1-t)\eta') \\
&= (t\eta + (1-t)\eta')\, s(1,\ t\eta + (1-t)\eta') \\
&\qquad + \left(1 - (t\eta + (1-t)\eta')\right) s(-1,\ t\eta + (1-t)\eta') \\
&= t\left[\eta\, s(1,\ t\eta + (1-t)\eta') + (1-\eta)\, s(-1,\ t\eta + (1-t)\eta')\right] \\
&\qquad + (1-t)\left[\eta'\, s(1,\ \text{——}) + (1-\eta')\, s(-1,\ \text{——})\right]. \\
&= t\, L(\eta,\ \text{——}) + (1-t)\, L(\eta',\ \text{——}). \\
&> t\, L(\eta, \eta) + (1-t)\, L(\eta', \eta') \\
&= -t\,\Omega(\eta) - (1-t)\,\Omega(\eta')
\end{aligned}
$$

□

[⇐] S: strictly proper ではない と仮定. i.e.

$\exists \eta \neq \eta'$ s.t. $\Omega(\eta) = -L(\eta, \eta) = -L(\eta, \eta') = \Omega(\eta') + g' \cdot (\eta - \eta')$

$g' \in \partial\Omega(\eta')$

このとき Ω が狭義凸でない ことを示す:

任意に $t \in (0,1)$ をとり.

$$\begin{aligned} -L(t\eta + (1-t)\eta',\ t\eta + (1-t)\eta') &= \Omega(t\eta + (1-t)\eta') \\ &\leq t\Omega(\eta) + (1-t)\Omega(\eta') \\ &= \Omega(\eta') + t g'(\eta - \eta'). \\ &= \Omega(\eta') + g' \cdot (t\eta + (1-t)\eta' - \eta') \\ &= -L(t\eta + (1-t)\eta',\ \eta') \\ &\leq -L(t\eta + (1-t)\eta',\ t\eta + (1-t)\eta') \end{aligned}$$

$\rightsquigarrow \Omega(t\eta + (1-t)\eta') = t\Omega(\eta) + (1-t)\Omega(\eta')$ ⌟

- Proper composite loss の導入 $\phi(y, \alpha) := S(y, \psi^{-1}(\alpha))$

  $\alpha \in \mathbb{R}$, $\psi : [0,1] \to \mathbb{R}$, $\hat{\eta} \mapsto \alpha$

ERM による学習に用いられる.

$$\min_{f : \mathcal{X} \to \mathbb{R}} \mathbb{E}\left[\phi(Y, f(X))\right] = \mathbb{E}\left[S(Y, \psi^{-1} \circ f(X))\right].$$

S: given のとき

Q. 「どのような ψ を選ぶと ϕ は convex になるか?

A. Canonical link $\psi(\hat{\eta}) := S(-1, \hat{\eta}) - S(1, \hat{\eta})$ を選ぶ

Prop. 3.2 s を strictly proper とし, $\psi(\hat{\eta}) := s(-1,\hat{\eta}) - s(1,\hat{\eta})$ とする.
ψ が $(0,1)$ で可逆であるとき, $\phi(\alpha) := s(1,\psi^{-1}(\alpha))$ は convex on $\alpha \in \psi((0,1))$

Proof 簡単のため, L が C^2-級である場合のみ示す.

composite loss $\phi(\alpha) = s(1,\psi^{-1}(\alpha))$ に対して, $\alpha = \psi(\hat{\eta})$ と書く.

Thm 3.1 より $\phi(\alpha) = L(1,\psi^{-1}(\alpha)) = L(\hat{\eta}) + L'(\hat{\eta})\cdot(1-\hat{\eta})$.

Thm 3.1 の証明中で $s(-1,\hat{\eta}) - s(1,\hat{\eta}) = -L'(\hat{\eta})$ が確認できるので,

$\psi(\hat{\eta}) = -L'(\hat{\eta})$ なら $L'(\hat{\eta}) = -\alpha$.

$\rightsquigarrow$ $\phi(\alpha) = L(\hat{\eta}) - \alpha(1-\hat{\eta})$.

$$\phi'(\alpha) = \underbrace{L'(\hat{\eta})}_{=-\alpha}\cdot\frac{d\hat{\eta}}{d\alpha} - (1-\hat{\eta}) + \alpha\frac{d\hat{\eta}}{d\alpha} = \hat{\eta} - 1$$

$$\phi''(\alpha) = \frac{d\hat{\eta}}{d\alpha} = \frac{1}{\psi'(\hat{\eta})} = -\frac{1}{L''(\hat{\eta})} \geq 0$$

(逆関数定理. / s : strictly proper.)

- **Remark** : canonical link は ϕ が convex になる唯一の link とは限らない.
- **Example 1** log score $s(1,\hat{\eta}) = -\log\hat{\eta}$, $s(-1,\hat{\eta}) = -\log(1-\hat{\eta})$.

条件つきリスク $L(\eta,\hat{\eta}) = -\eta\log\hat{\eta} - (1-\eta)\log(1-\hat{\eta})$ (交差エントロピー)

canonical link $\psi(\hat{\eta}) = s(-1,\hat{\eta}) - s(1,\hat{\eta}) = \log\left(\frac{\hat{\eta}}{1-\hat{\eta}}\right)$ (logit link)

Bayes リスク $L(\eta) = L(\eta,\eta) = -\eta\log\eta - (1-\eta)\log(1-\eta)$ を用いると,

$L'(\eta) = -\log\eta - 1 + \log(1-\eta) + 1 = -\log\left(\frac{\eta}{1-\eta}\right) = -\psi(\hat{\eta})$ が現れる.

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

(logistic loss)

- **Example 2.** Brier score $s(1,\hat{\eta}) = (1-\hat{\eta})^2$, $s(-1,\hat{\eta}) = \hat{\eta}^2$.

  条件つきリスク $L(\eta,\hat{\eta}) = \eta(1-\hat{\eta})^2 + (1-\eta)\hat{\eta}^2$.

  Bayes リスクに関して: $L(\eta,\cdot)$ の最小化を考える.

  $$\frac{\partial L}{\partial \hat{\eta}} = 2\eta(\hat{\eta}-1) + 2(1-\eta)\hat{\eta} = 2(\eta-\hat{\eta}) = 0 \to \hat{\eta} = \eta$$

  $\rightsquigarrow \underline{L}(\eta) = L(\eta,\eta) = \eta(1-\eta)^2 + (1-\eta)\eta^2 = \eta(1-\eta)$. (Gini index)

  $\rightsquigarrow$ s は strictly proper.

  canonical link $\psi(\hat{\eta}) = s(-1,\hat{\eta}) - s(1,\hat{\eta}) = 2\hat{\eta} - 1$

  ($\underline{L}'(\hat{\eta}) = 1-2\hat{\eta}$ からも確認可能)

  composite loss $\phi(\alpha) = \left(1 - \frac{1+\alpha}{2}\right)^2 = \frac{1}{4}(1-\alpha)^2$.

  二乗損失.

- **Example 3.** Power score $s(1,\hat{\eta}) = (1-\hat{\eta})^r$, $s(-1,\hat{\eta}) = \hat{\eta}^r$ $(r \geq 1)$

  条件つきリスク $L(\eta,\hat{\eta}) = \eta(1-\hat{\eta})^r + (1-\eta)\hat{\eta}^r$,

  L の最小化を考える: $\frac{\partial L}{\partial \hat{\eta}} = -r\eta(1-\hat{\eta})^{r-1} + r(1-\eta)\hat{\eta}^{r-1} = 0$.

  $\rightsquigarrow \eta(1-\hat{\eta})^{r-1} = (1-\eta)\hat{\eta}^{r-1}$

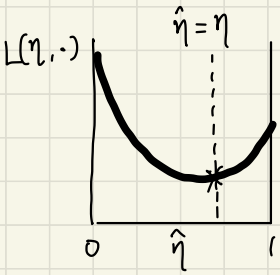
  ここで $\hat{\eta} = \eta$ を代入すると $\eta^{r-2} = (1-\eta)^{r-2}$

  これは $r=2$ 以外では不成立. したがって proper であるのは $r=2$ のみ.

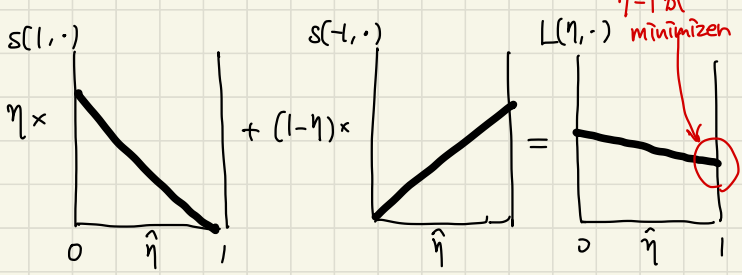
  Brier.

図で解釈.

$r=2$ のとき.



$r=1$ のとき.



- Example 4. $s(1,\hat{\eta}) = \sqrt{\frac{1-\hat{\eta}}{\hat{\eta}}}$, $s(-1,\hat{\eta}) = \sqrt{\frac{\hat{\eta}}{1-\hat{\eta}}}$

  ⇝ strictly proper, $\underline{L}(\eta) = 2\sqrt{\eta(1-\eta)}$

  ⇝ canonical link $\psi(\hat{\eta}) = \sqrt{\frac{\hat{\eta}}{1-\hat{\eta}}} - \sqrt{\frac{1-\hat{\eta}}{\hat{\eta}}}$.

  composite loss $\phi(\alpha) = \frac{\sqrt{\alpha^2+4} - \alpha}{2}$ (convex)

しかし, $s(1,\hat{\eta})$ に $\hat{\eta} = \frac{1}{1+e^{-\alpha}}$ を 逆リンク関数 として代入.

⇝ $s(1, \frac{1}{1+e^{-\alpha}}) = e^{-\alpha/2}$ (convex).

つまり canonical link は ϕ を convex にする唯一のリンクとは限らない.