Non-Paramatic Statistic Sheet 2

Osman Ceylan, Jiahui Wang, Zhuoyao Zeng

11. November 2020

Exercise 1.1

Let $(X, \mathcal{A})$ be a measurable space, $Y := \{-1, 1\}$, and P be a distribution on $X \times Y$. Define $\eta : X \to [0, 1]$ by $\eta(x) := P(\{1\}|x)$ for all $x \in X$.

i) Use η to determine the Bayes risk and all Bayes decision functions for the binary classification loss $L_{\text{class}} : Y \to [0, \infty)$.

Basic Assumption:

We have regular conditional probability $P(\cdot|x)$ met the following conditions:

1) $P(\cdot|x) : \{-1, 1\} \to [0, 1]$ is a probability measure
2) $x \mapsto P(B|x)$ is measurable for $B = \{1\}$ or $B = \{-1\}$
3) $P(A \times B) = \int_A P(B|x)\, \mathrm{d}P_X$

Since the prediction $t \in \{0, 1, -1\}$, we define $\tilde{t} := \operatorname{sign}(t)$ with $\operatorname{sign}(0) = 1$ to simplify the expression.

Now we look for our $C^*_{L,p}(x) := \inf_{\tilde{t} \in \{-1,1\}} \{ \int_Y L_{\text{class}}(y, \tilde{t})\, P(\mathrm{d}y|x) \}$.

$$\int_Y L_{\text{class}}(y, \tilde{t})\, P(\mathrm{d}y|x) = \int_Y \mathbb{1}_{(0,\infty]}(y \cdot \tilde{t}) P(\mathrm{d}y|x)$$
$$= 1 \cdot P(y \cdot \tilde{t} \leq 0|x) = \begin{cases} P(\{1\}|x) = \eta(x), & \tilde{t}(x) = -1 \\ P(\{-1\}|x) = 1 - \eta(x), & \tilde{t}(x) = 1 \end{cases}.$$

To minimize our integral in $\tilde{t}$ for a given x, we just need to compare $\eta(x)$ with $1 - \eta(x)$. More precisely, we make the following choice:

When $\eta(x) \geq 1 - \eta(x) \;\Leftrightarrow\; \eta(x) \geq \frac{1}{2}$, we choose $\tilde{t}(x) = 1$,

When $\eta(x) < 1 - \eta(x) \;\Leftrightarrow\; \eta(x) < \frac{1}{2}$, we choose $\tilde{t}(x) = -1$

Now consider all $x \in X$. The target function t^*, according to the algorithm above, should be:

$$t^*(x) = \begin{cases} 1, & \eta(x) \geq \frac{1}{2} \\ -1, & \eta(x) < \frac{1}{2} \end{cases} \quad = \mathbb{1}_{\{\eta \geq \frac{1}{2}\}} - \mathbb{1}_{\{\eta < \frac{1}{2}\}}.$$

$\Rightarrow \quad C^*_{L,p}(x) = \int_Y \mathbb{1}_{(0,\infty]}(y \cdot t^*(x)) P(\mathrm{d}y|x) = 1 \cdot P(y \cdot t^*(x) \leq 0|x) = \min\{\eta(x), 1 - \eta(x)\}$ P_X-almost surely.

According to (1.2.8), all functions which satisfy the equation above is a Bayes decision function.

Now we compute the Bayes Risk:
$R^*_{L,P} = \int_X C^*_{L,p}(x)\mathrm{d}P_X = \int_{\{\eta(x)\geq\frac{1}{2}\}}(1-\eta(x))\mathrm{d}P_X + \int_{\{\eta(x)<\frac{1}{2}\}}\eta(x)\mathrm{d}P_X.$

ii) Given a so-called weight parameter $\alpha \in (0,1)$ and consider the α-weighted binary classification loss $L_{\text{class},\alpha} : Y \times \mathbb{R} \to [0,\infty)$ defined by:

$$L_{\text{class},\alpha} = \begin{cases} 1-\alpha, & y = 1 \wedge t < 0 \\ \alpha, & y = -1 \wedge t \geq 0 \\ 0, & \text{otherwise} \end{cases} \cdot$$

Determine the corresponding Bayes risk and Bayes decision functions and compare your findings to part i).

Analogously to part i), we want to minimize the average loss in t according to $P(\mathrm{d}y|x)$ for every given $x \in X$. In the second case,

$$\int_Y L_{\text{class},\alpha}(y,\tilde{t})\, P(\mathrm{d}y|x) = \begin{cases} (1-\alpha)\cdot P(\{1\}|x) = (1-\alpha)\cdot\eta(x), & \tilde{t}(x) = -1 \\ \alpha\cdot P(\{-1\}|x) = \alpha\cdot(1-\eta(x)), & \tilde{t}(x) = 1 \end{cases}.$$

And we make the following choice:
When $(1-\alpha)\eta(x) \geq \alpha(1-\eta(x)) \Leftrightarrow \eta(x) \geq \alpha$, we choose $\tilde{t}(x) = 1$,
When $(1-\alpha)\eta(x) < \alpha(1-\eta(x)) \Leftrightarrow \eta(x) < \alpha$, we choose $\tilde{t}(x) = -1$,
Now consider all $x \in X$. The target function t^*, according to the algorithm above, should be:

$$t^*(x) = \begin{cases} 1, & \eta(x) \geq \alpha \\ -1, & \eta(x) < \alpha \end{cases} \;=\; \mathbb{1}_{\{\eta\geq\alpha\}} - \mathbb{1}_{\{\eta<\alpha\}}.$$

$\Rightarrow \quad C^*_{L,p}(x) = \int_Y L_{\text{class},\alpha}(y,t^*)\, P(\mathrm{d}y|x) \;=\; \min\{(1-\alpha)\eta(x), \alpha(1-\eta(x))\}$ P_X-almost surely.
According to (1.2.8), all functions which satisfy the equation of integral above is a Bayes decision function.

Now we compute the Bayes Risk:
$R^*_{L,P} = \int_X C^*_{L,p}(x)\mathrm{d}P_X = \int_{\{\eta(x)\geq\alpha\}}\alpha(1-\eta(x))\mathrm{d}P_X + \int_{\{\eta(x)<\alpha\}}(1-\alpha)\eta(x)\mathrm{d}P_X.$