

Non-Paramatic Statistic Sheet 2

Osman Ceylan, Jiahui Wang, Zhuoyao Zeng

13. 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 $\eta : X \rightarrow [0, 1]$ 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 $L_{\text{class}} : Y \rightarrow [0, \infty)$.

Solution:

Basic assumptions:

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

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

Since the prediction $t \in \{0, 1, -1\}$ or actually should it be $t \in \mathbb{R}$, we define $\tilde{t} := \text{sign}(t)$ with $\text{sign}(0) = 1$ to simplify the expression.

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

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

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 choices for an $x \in X$:

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:

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

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

According to (1.2.9), all functions satisfying the equation above are Bayes decision functions P_X -almost surely..

Now we compute the Bayes Risk:

$$R_{L,P}^* = \int_X C_{L,p}^*(x) dP_X(x) = \int_{\{\eta(x) \geq \frac{1}{2}\}} (1 - \eta(x)) dP_X(x) + \int_{\{\eta(x) < \frac{1}{2}\}} \eta(x) dP_X(x).$$

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

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

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

Solution:

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

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

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:

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

$$\Rightarrow C_{L,p}^*(x) = \int_Y L_{\text{class},\alpha}(y, t^*) P(dy|x) = \min\{(1 - \alpha)\eta(x), \alpha(1 - \eta(x))\} \quad P_X\text{-almost surely.}$$

According to (1.2.8) all functions satisfying the equation of integral above are Bayes decision functions.

Now we compute the Bayes Risk:

$$R_{L,P}^* = \int_X C_{L,p}^*(x) dP_X = \int_{\{\eta(x) \geq \alpha\}} \alpha(1 - \eta(x)) dP_X + \int_{\{\eta(x) < \alpha\}} (1 - \alpha)\eta(x) dP_X.$$

Exercise 1.4

Let (X, \mathcal{A}) be a measurable space, μ be a σ -finite measure on X , and $h_{-1}, h_1 : X \rightarrow [0, \infty)$ be measurable with

$$\int_X h_{-1} d\mu = \int_X h_1 d\mu = 1. \quad (1.5.1)$$

For a given $p \in [0, 1]$ and $Y := \{-1, 1\}$ consider the distribution P on $X \times Y$ that is uniquely determined by

$$P(\{y\} \times A) := \begin{cases} (1-p) \int_A h_{-1} d\mu, & y = -1 \\ p \int_A h_1 d\mu, & y = 1 \end{cases}. \quad (1.5.2)$$

for all $y \in Y$ and $A \in \mathcal{A}$.

i) Determine the marginal distribution P_X and the function $\eta : X \rightarrow [0, 1]$ that is given by $\eta(x) := P(\{1\} | x)$ for all $x \in X$.

Solution:

Consider an arbitrary $A \in \mathcal{A}$.

Observe that $Y \times A = \{-1, 1\} \times A = \{-1\} \times A \cup \{1\} \times A$.

With σ -additivity of measure we have:

$$\begin{aligned} P_X(A) &= P(Y \times A) = P(\{-1\} \times A) + P(\{1\} \times A) \\ &= (1-p) \int_A h_{-1} d\mu + p \int_A h_1 d\mu = \int_A (1-p)h_{-1} + p h_1 d\mu. \end{aligned}$$

Also observe that $f := (1-p)h_{-1} + p h_1$ is a positive function. (#)

$\Rightarrow f$ is the density of P_X with respect to μ . (*)

For η we consider the quantity $P(\{1\} \times A)$ along with the equality 1.2.3 from lecture slides:

$$P(\{1\} \times A) \stackrel{1.2.3}{=} \int_A P(1|x) dP_X(x) \stackrel{(*)}{=} \int_A P(1|x) f(x) d\mu(x).$$

Because of (1.5.2) and $\eta = P(\{1\} | x)$ we then have: $\int_A p h_1 d\mu = \int_A \eta f d\mu$.

Particularly it holds that $\int_X p h_1 d\mu = \int_X \eta f d\mu$.

\Rightarrow For μ -almost every $x \in X$ we have $p h_1 = \eta f$.

$\stackrel{(\#)}{\Rightarrow}$ For μ -almost every $x \in X$ it holds that $\eta = \frac{p h_1}{f} = \frac{p h_1}{(1-p)h_{-1} + p h_1}$.

(*) says that $P_X \ll \mu \Rightarrow$ For P_X -almost every $x \in X$ it holds that $\eta = \frac{p h_1}{(1-p)h_{-1} + p h_1}$.

ii) Show that for all distributions P on $X \times Y$ there exists μ, p , and measurable $h_{-1}, h : X \rightarrow [0, \infty)$ with (1.5.1) such that (1.5.2) holds.

Solution:

$Y = \{-1, 1\}$ is a closed subset of \mathbb{R}

\Rightarrow regular conditional probability $P(\cdot | \cdot)$ with respect to P exists.

Define $\eta(x) := P(\{1\} | x)$ which is measurable per Definition of $P(\cdot | \cdot)$.

Consider $p := P(\{1\} \times X) \in [0, 1]$ as P is a probability measure.

We firstly consider $p \in \{0, 1\}$, and without loss of generality, we consider $p = 1$.

In this case, we can choose $\mu := P_X$, $h_1 = h_{-1} := 1$ the constant function on X .

It is then easy to calculate, that (1.5.1) and (1.5.2) will be satisfied under this choice.

Now suppose $p \in (0, 1)$.

We propose that with the following quantities (1.5.1) and (1.5.2) can be satisfied:

$$h_1 := (1 - p)\eta, \quad h_{-1} := p(1 - \eta)$$

$$\mu : \mathcal{A} \rightarrow [0, \infty), A \mapsto \frac{P(Y \times A)}{p(1 - p)} = \frac{1}{p(1 - p)} P_X(A).$$

Notice that: h_1, h_{-1} are measurable since η is measurable, and it holds that $p(1 - p)\mu = P_X$.

Now we verify (1.5.1):

$$\int_X h_1 d\mu = \int_X p(1 - \eta) d\mu = \int_X \frac{p(1 - \eta)}{p(1 - p)} dP_X = \frac{1}{1 - p} \int_X 1 - P(\{1\}|x) dP_X = \frac{1}{1 - p} (1 - p) = 1.$$

Similarly it can be calculated that $1 = \int_X p(1 - \eta) d\mu = \int_X h_{-1} d\mu$.

(1.5.2) should also be verified:

$$\begin{aligned} P(\{1\} \times A) &= \int_A P(\{1\}|x) dP_X \\ &= \int_A \eta dP_X \\ &= \int_A \eta p(1 - p) d\mu \\ &= p \int_A (1 - p) \eta d\mu \\ &= p \int_A h_1 d\mu. \end{aligned} \quad \begin{aligned} P(\{-1\} \times A) &= \int_A P(\{-1\}|x) dP_X \\ &= \int_A (1 - P(\{1\}|x)) dP_X \\ &= \int_A (1 - \eta) p(1 - p) d\mu \\ &= (1 - p) \int_A h_{-1} d\mu. \end{aligned}$$

iii) Find an interpretation of your findings in the spirit of Bayes's theorem.

Solution:

Let P an arbitrary distribution on $X \times Y$ and $x_0 \in X$ an arbitrary element.

$(A_n)_{n \in \mathbb{N}}$ is a sequence of monotonous decreasing subsets in X , with $x_0 \in A_n$ for all $n \in \mathbb{N}$.

According to Bayes's theorem we'll have the equation:

$$P(\{y = 1\} | x \in A_n) = \frac{P(\{y = 1\} \cap \{x \in A_n\})}{P(\{x \in A_n\})} \stackrel{def.}{=} \frac{P(\{1\} \times A_n)}{P_X(A_n)}$$

Now, we insert (1.5.1) and (1.5.2) into this equation and receive:

$$P(\{y = 1\} | x \in A_n) = \frac{p \int_A h_1 d\mu}{p \int_A h_1 d\mu + (1 - p) \int_A h_{-1} d\mu}$$

By taking the limit of A_n to $\{x_0\}$, we will have (stimmt das???)

$$\eta(x_0) = P(\{1\}|x_0) = \frac{p h_1(x_0)}{(1 - p) h_1(x_0) + p h_1(x_0)}.$$

It is also the relationship we find in part i).