Logistic regression and classification problems

Teaching team of STA314

Department of Statistical Sciences
University of Toronto

Monday October 7th, 2024

Classification problems

Problem set 3 Q2

Consider the classification problem with the label of Y belong to $\mathcal{C} := \{1, 2, \ldots, K\}$ and any realization x of $X \in \mathbb{R}^p$. The Bayes classifier f^* at $X = x$ is defined as

$$f^*(x) := \underset{f(x)\in\mathcal{C}}{\operatorname{argmin}} \ \mathbb{E}\Big[1\{Y \neq f(X)\} \,|\, X = x\Big]$$

1. Prove that

$$f^*(x) = \underset{k\in\mathcal{C}}{\operatorname{argmax}} \ \mathbb{P}(Y = k \mid X = x).$$

2. Prove that the Bayes error at $X = x$ equals to

$$1 - \max_{k\in\mathcal{C}} \ \mathbb{P}(Y = k \mid X = x).$$

Problem set 3 - Q2

1. Prove that
   $f^*(x) = \text{argmax}_{k \in \mathcal{C}} \ \mathbb{P}(Y = k \mid X = x)$.

Proof.

$$\begin{aligned}
\mathbb{P}(Y = f(X) \mid X = x) &= \sum_{k \in \mathcal{C}} \mathbb{P}(Y = f(X) = k \mid X = x) \\
&= \sum_{k \in \mathcal{C}} \mathbb{P}(Y = k \mid X = x) \, 1\{f(x) = k\} \\
&\leq \max_{k \in \mathcal{C}} \mathbb{P}(Y = k \mid X = x) .
\end{aligned}$$

Moreover, the maximal value is achieved when $f(x) = k^*$ for

$$k^* = \underset{k \in \mathcal{C}}{\text{argmax}} \, \mathbb{P}(Y = k \mid X = x) .$$

This proves the first claim. □

Problem set 3 Q2

2. Prove that the Bayes error at $X = x$ equals to

$$1 - \max_{k \in \mathcal{C}} \mathbb{P}\left(Y = k \mid X = x\right).$$

Proof.

By similar reasoning,

$$\begin{aligned}
\mathbb{P}(Y \neq f^*(X) \mid X = x) &= 1 - \mathbb{P}(Y = f^*(X) \mid X = x) \\
&= 1 - \sum_{k \in \mathcal{C}} \mathbb{P}(Y = k \mid X = x) 1\{f^*(x) = k\} \\
&= 1 - \max_{k \in \mathcal{C}} \mathbb{P}(Y = k \mid X = x).
\end{aligned}$$

$\square$

Problem set 3 Q2

3. Consider that $K = 3$. For a fixed x_0, assume that

$$\mathbb{P}(Y = 1 \mid X = x_0) = 0.5$$
$$\mathbb{P}(Y = 2 \mid X = x_0) = 0.3$$
$$\mathbb{P}(Y = 3 \mid X = x_0) = 0.2.$$

State the Bayes classifier at $X = x_0$ and compute its error at $X = x_0$.

Solution.

As a reminder, the Bayes classifier f^* at $X = x$ is defined as

$$f^*(x) := \underset{f(x)\in\mathcal{C}}{\text{argmin}}\ \mathbb{E}\Big[1\{Y \neq f(X)\} \,\big|\, X = x\Big]$$

We can see that $f^*(x_0) = 1$ since that's the label with highest probability.
Using the solution from 2 the Bayes error at $X = x_0$ is 0.5

□

Problem set 3 Q2

4. Consider a naive classifier $\hat{f}$, called random guessing, which randomly picks one label from $\mathcal{C} = \{1, 2, 3\}$ with equal probability. Compare its expected error rate at $X = x_0$ with the Bayes error from part 3.

Solution (1/2).

The expected error rate of $\hat{f}$ at x_0 is

$$\begin{aligned}
\mathbb{E}\Big[1\{\hat{f}(X) \neq Y\} \mid X = x_0\Big] &= \mathbb{P}(\hat{f}(X) \neq Y \mid X = x_0) \\
&= 1 - \mathbb{P}(\hat{f}(X) = Y \mid X = x_0) \\
&= 1 - \sum_{c \in \mathcal{C}} \mathbb{P}(Y = c \,|\, X = x_0)\mathbb{P}(\hat{f}(X) = c | X = x_0) \\
&= 1 - \sum_{c \in \mathcal{C}} \mathbb{P}(Y = c \mid X = x_0)\mathbb{P}(\hat{f}(X) = c)
\end{aligned}$$

Continues in next slide □

Problem set 3 Q2

Solution (2/2).

The expected error rate of $\hat{f}$ at x_0 is

$$\begin{aligned}
\mathbb{E}\Big[1\{\hat{f}(X) \neq Y\} \mid X = x_0\Big] &= 1 - \sum_{c \in \mathcal{C}} \mathbb{P}(Y = c \mid X = x_0)\mathbb{P}(\hat{f}(X) = c) \\
&= 1 - \frac{1}{3} \sum_{c \in \mathcal{C}} \mathbb{P}(Y = c \mid X = x_0) \\
&= 1 - \frac{1}{3}[0.5 + 0.3 + 0.2] = \frac{2}{3},
\end{aligned}$$

which is greater than the Bayes error. □

Problem set 3 Q3

We will prove that a random guessing classifier for binary classification has the area under the curve (AUC) equal to 1/2.
Suppose $Y \in \{0, 1\}$ with $Y = 1$ meaning *true*, and *false* otherwise.
Consider the following random guessing classifier at any $X = x$

$$\hat{f}(x) = \begin{cases} 1, & \text{with prob. equal to } p \\ 0, & \text{with prob. equal to } 1 - p \end{cases}$$

with any p chosen from $[0, 1]$.

Problem set 3 Q3

1. Prove that the expected AUC of this random classifier with p varying within $[0, 1]$ is $1/2$.

Proof.

The expected FPR is

$$\mathbb{P}(\hat{f}(X) = 1 \mid Y = 0) = \mathbb{P}(\hat{f}(X) = 1 = p$$

while the expected FNR is

$$\mathbb{P}(\hat{f}(X) = 0 \mid Y = 1) = \mathbb{P}(\hat{f}(X) = 0) = 1 - p.$$

Note that the expected TPR is $1 - \mathbb{P}(\hat{f}(X) = 0 \mid Y = 1) = p$. Therefore, the AUC of $\hat{f}$ is

$$\int_0^1 p dp = \frac{1}{2}.$$

□

Problem set 3 Q3

2. Let $\eta(x) := \mathbb{P}(Y = 1 \mid X = x)$ for any x. Write the expected error rate of $\hat{f}$ at $X = x$ in terms of $\eta(x)$ and p.

Proof.

The expected error rate at $X = x$ equals to

$$
\begin{aligned}
&\mathbb{P}(Y \neq \hat{f}(x) \mid X = x) \\
&= \mathbb{P}(Y = 1, \hat{f}(x) = 0 \mid X = x) + \mathbb{P}(Y = 0, \hat{f}(x) = 1 \mid X = x) \\
&= \mathbb{P}(Y = 1 \mid X = x)\mathbb{P}(\hat{f}(x) = 0 \mid X = x) \\
&\qquad + \mathbb{P}(Y = 0 \mid X = x)\mathbb{P}(\hat{f}(x) = 1 \mid X = x) \\
&= \mathbb{P}(Y = 1 \mid X = x)\mathbb{P}(\hat{f}(x) = 0) + \mathbb{P}(Y = 0 \mid X = x)\mathbb{P}(\hat{f}(x) = 1) \\
&= \eta(x)(1 - p) + (1 - \eta(x))p.
\end{aligned}
$$

□

Problem set 3 Q3

3. If you have the flexibility of choosing p in your expression of part 2, which choice minimizes the expected error rate of $\hat{f}$ at $X = x$? Is the resulting classifier equivalent to the Bayes classifier? State your explanation.

Proof.

We choose

$$p = \begin{cases} 1 & \text{if } \eta(x) \geq 1/2 \\ 0 & \text{otherwise} \end{cases}$$

The resulting classifier only equals to the Bayes classifier at this point but not in general. Indeed, to match the Bayes classifier for all x, we need to allow $p = p(x)$ being a function that depends on x. □