Statistical Foundations of Learning - Sample Problems 2

CIT4230004 (Summer Semester 2024)

Sample Problem 2.1: Bayes Risk for K classes

Consider a classification problem of K classes, where we define $\eta_k(x) = \mathbb{P}(Y = k \mid X = x)$ for all $x \in X$, $k \in [K]$. Assume that i.i.d. training sample pairs $S = \{(x_i, y_i)\}_{i=1}^m$ are drawn from some joint distribution D. Similarly, we define a test example as $(x, y) \sim D$.

In a supervised learning setting, the goal is to find a classification rule $\hat{h}(\cdot)$ such that the expected risk over S and an unseen test example is small.

(a) Write out the risk of a classifier h for this classification problem.

The risk of a classifier h is defined as the expected loss over the distribution D:

$$R(h) = \mathbb{E}_{(X,Y)\sim D}[\ell(h(X), Y)]$$

where ℓ is the loss function (e.g., 0-1 loss).

(b) Write out the Bayes risk.

The Bayes risk is the minimum possible risk, achieved by the Bayes classifier h^*:

$$R(h^*) = \min_h R(h)$$

$$R(h^*) = \mathbb{E}_X \left[\min_{k\in[K]} (1 - \eta_k(X)) \right]$$

where $\eta_k(x) = \mathbb{P}(Y = k \mid X = x)$.

Sample Problem 2.2: Bayes Risk for Uniform Features X

Let $X = \mathbb{R}$ and $Y = \{\pm 1\}$. Define a distribution D such that $(x, y) \sim D$ implies:

$$x \sim \text{Uniform}[0, 3],$$

$$\mathbb{P}(Y = 1 \mid x) = \begin{cases} \frac{3}{4} & \text{if } x \in (1, 2), \\ \frac{1}{4} & \text{if } x \in [0, 1] \cup [2, 3]. \end{cases}$$

(a) Compute the Bayes risk for the problem.

The Bayes classifier h^* assigns the most probable class at each x:

$$h^*(x) = \begin{cases} 1 & \text{if } x \in (1,2), \\ -1 & \text{if } x \in [0,1] \cup [2,3]. \end{cases}$$

The Bayes risk is:

$$R(h^*) = \mathbb{E}\left[\min_{k\in\{\pm 1\}}(1-\eta_k(x))\right]$$

For $x \in (1,2)$:

$$\min(1-\eta_1(x), 1-\eta_{-1}(x)) = 1 - \frac{3}{4} = \frac{1}{4}$$

For $x \in [0,1] \cup [2,3]$:

$$\min(1-\eta_1(x), 1-\eta_{-1}(x)) = 1 - \frac{1}{4} = \frac{3}{4}$$

The Bayes risk is:

$$R(h^*) = \int_0^3 \min(1-\eta_1(x), 1-\eta_{-1}(x)) f_X(x)\, dx$$

where $f_X(x) = \frac{1}{3}$.

Thus,

$$R(h^*) = \frac{1}{3}\left(\int_0^1 \frac{3}{4}\, dx + \int_1^2 \frac{1}{4}\, dx + \int_2^3 \frac{3}{4}\, dx\right)$$

$$R(h^*) = \frac{1}{3}\left(\frac{3}{4}\cdot 1 + \frac{1}{4}\cdot 1 + \frac{3}{4}\cdot 1\right)$$

$$R(h^*) = \frac{1}{3}\left(\frac{3}{4} + \frac{1}{4} + \frac{3}{4}\right)$$

$$R(h^*) = \frac{1}{3}\cdot\frac{7}{4} = \frac{7}{12}$$

(b) Given $t \in \mathbb{R}$, $b \in \{\pm 1\}$, define a classifier

$$h_{t,b}(x) = \begin{cases} b & \text{if } x \le t, \\ -b & \text{if } x > t. \end{cases}$$

Compute the risk of $h_{t,b}$ in terms of t, b.
The risk is:

$$R(h_{t,b}) = \mathbb{E}\left[\ell(h_{t,b}(X), Y)\right]$$

For $b = 1$:

$$R(h_{t,1}) = \int_0^t \mathbb{P}(Y=-1 \mid x) f_X(x)\, dx + \int_t^3 \mathbb{P}(Y=1 \mid x) f_X(x)\, dx$$

For $b = -1$:

$$R(h_{t,-1}) = \int_0^t \mathbb{P}(Y = 1 \mid x) f_X(x)\, dx + \int_t^3 \mathbb{P}(Y = -1 \mid x) f_X(x)\, dx$$

Using $f_X(x) = \frac{1}{3}$ and $\mathbb{P}(Y = 1 \mid x)$:

$$R(h_{t,1}) = \frac{1}{3} \left(\int_0^t (1 - \mathbb{P}(Y = 1 \mid x))\; dx + \int_t^3 \mathbb{P}(Y = 1 \mid x)\, dx \right)$$

$$R(h_{t,-1}) = \frac{1}{3} \left(\int_0^t \mathbb{P}(Y = 1 \mid x)\, dx + \int_t^3 (1 - \mathbb{P}(Y = 1 \mid x))\; dx \right)$$

(c) Which t, b achieves the minimum risk?
Evaluate t and b that minimize the risk: For $t \in [0, 1]$:

$$R(h_{t,1}) = \frac{1}{3} \left(\frac{3}{4} t + \frac{3}{4}(3 - t) \right) = \frac{1}{3} \cdot \frac{3}{4} \cdot 3 = \frac{3}{4}$$

For $t \in [1, 2]$:

$$R(h_{t,1}) = \frac{1}{3} \left(\frac{3}{4} t + \frac{1}{4}(2 - t) + \frac{3}{4}(3 - 2) \right)$$

$$= \frac{1}{3} \left(\frac{3}{4} t + \frac{1}{4}(2 - t) + \frac{3}{4} \right)$$

For $t \in [2, 3]$:

$$R(h_{t,1}) = \frac{1}{3} \left(\frac{3}{4} + \frac{1}{4}(t - 2) + \frac{1}{4}(3 - t) \right)$$

$$= \frac{1}{3} \left(\frac{3}{4} + \frac{1}{4} \right) = \frac{1}{3} \cdot 1 = \frac{1}{3}$$

Thus, the minimum risk is achieved for $t = 2$ and $b = 1$.
Therefore, the optimal classifier is:

$$h_{2,1}(x) = \begin{cases} 1 & \text{if } x \leq 2, \\ -1 & \text{if } x > 2. \end{cases}$$

Sample Problem 2.3: Convergence to Nearest Neighbours

Consider $X \subseteq \mathbb{R}^p$ and a continuous distribution D_X on X with probability density $f(x)$. Let $x_1, \ldots, x_m \sim$ i.i.d. D_X. Fix an integer $k \in \mathbb{N}$.

For a point $x^* \in X$, we denote by $B(x^*, \epsilon) \subseteq X$ the ball of points that have a distance of at most ϵ from x^*. Recall that we defined:

$$D_X(x^*; \epsilon) = \mathbb{P}_{x \sim D_X}(x \in B(x^*, \epsilon))$$

Let $x_{\pi_k(x^*)} \in \{x_1, \ldots, x_m\}$ denote the k-th nearest neighbour of x^*. The following steps prove that the second nearest neighbour $x_{\pi_2(x^*)} \to x^*$ in probability as $m \to \infty$.

(a) Let $N_\epsilon = |\{i : x_i \in B(x^*, \epsilon)\}|$. State the distribution of N_ϵ in terms of m and $D_X(x^*; \epsilon)$, and give an expression for the probability $\mathbb{P}(N_\epsilon < 2)$.

The number of points N_ϵ in $B(x^*, \epsilon)$ follows a binomial distribution:

$$N_\epsilon \sim \text{Binomial}(m, D_X(x^*; \epsilon))$$

The probability that there are fewer than 2 points in $B(x^*, \epsilon)$ is:

$$\mathbb{P}(N_\epsilon < 2) = \mathbb{P}(N_\epsilon = 0) + \mathbb{P}(N_\epsilon = 1)$$

Using the binomial probability formula:

$$\mathbb{P}(N_\epsilon = 0) = (1 - D_X(x^*; \epsilon))^m$$

$$\mathbb{P}(N_\epsilon = 1) = mD_X(x^*; \epsilon)(1 - D_X(x^*; \epsilon))^{m-1}$$

Therefore:

$$\mathbb{P}(N_\epsilon < 2) = (1 - D_X(x^*; \epsilon))^m + mD_X(x^*; \epsilon)(1 - D_X(x^*; \epsilon))^{m-1}$$

(b) Use part (a) to bound $\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon))$ and show that if $D_X(x^*; \epsilon) > 0$, then $\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)) \to 0$ as $m \to \infty$.

To bound $\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon))$, we use the fact that if $N_\epsilon \geq 2$, then the second nearest neighbour $x_{\pi_2(x^*)}$ must be within $B(x^*, \epsilon)$.

Thus:

$$\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)) \leq \mathbb{P}(N_\epsilon < 2)$$

Using the expression from part (a):

$$\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)) \leq (1 - D_X(x^*; \epsilon))^m + mD_X(x^*; \epsilon)(1 - D_X(x^*; \epsilon))^{m-1}$$

As $m \to \infty$, if $D_X(x^*; \epsilon) > 0$:

$$(1 - D_X(x^*; \epsilon))^m \to 0$$

$$mD_X(x^*; \epsilon)(1 - D_X(x^*; \epsilon))^{m-1} \to 0$$

Therefore:

$$\mathbb{P}(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)) \to 0$$

(c) Give an expression for $D_X(x^*; \epsilon)$ and, assuming $f(x) \geq f_{\min} > 0$ for all $x \in X$, show that $D_X(x^*; \epsilon) > 0$ for every $\epsilon > 0$.

The probability $D_X(x^*; \epsilon)$ is:

$$D_X(x^*; \epsilon) = \int_{B(x^*, \epsilon)} f(u)\, du$$

Assuming $f(x) \geq f_{\min} > 0$ for all $x \in X$, we have:

$$D_X(x^*; \epsilon) \geq f_{\min} \int_{B(x^*, \epsilon)} du = f_{\min} \cdot \text{Volume}(B(x^*, \epsilon))$$

Since the volume of the ball $B(x^*, \epsilon)$ is positive for any $\epsilon > 0$, it follows that:

$$D_X(x^*; \epsilon) > 0$$

Thus, $D_X(x^*; \epsilon) > 0$ for every $\epsilon > 0$.