Sample Problems 2

To be discussed on 10.05.2024

Sample Problem 2.1: Bayes Risk for K classes

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

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.

(b) Write out the Bayes risk.

Hint: The solution should be given in terms of an expectation over x.

Sample Problem 2.2: Bayes Risk for Uniform Features X

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

$$x \sim \text{Uniform}[0, 3], \qquad \text{and} \qquad P_{\mathcal{Y}|\mathcal{X}}(y = 1|x) = \begin{cases} \frac{3}{4} & \text{if } x \in (1, 2), \\ \\ \frac{1}{4} & \text{if } x \in [0, 1] \cup [2, 3]. \end{cases}$$

(a) Compute the Bayes risk for the problem.

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

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

The function $h_{t,b}$ is often referred to as a decision stump. Compute the risk of $h_{t,b}$ in terms of t, b.

(c) Which t, b achieves the minimum risk?

Sample Problem 2.3: Convergence to Nearest Neighbours

Consider $\mathcal{X} \subset \mathbb{R}^p$ and a continuous distribution $\mathcal{D}_{\mathcal{X}}$ on $\mathcal{X}$ with probability density $f(x)$. Let $x_1, \ldots, x_m \sim_{iid} \mathcal{D}_{\mathcal{X}}$. Fix an integer $k \in \mathbb{N}$.

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

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

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

(b) Use part (a) to bound $\mathbb{P}\left(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)\right)$ and show that if $\mathcal{D}_{\mathcal{X}}(x^*; \epsilon) > 0$, then

$$\mathbb{P}\left(x_{\pi_2(x^*)} \notin B(x^*, \epsilon)\right) \to 0 \quad \text{as } m \to \infty.$$

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