Assignment 1

Due: 14.05.2024, 15:00

Points: 17

The solutions have to be handed in via Moodle. We do not accept late submissions.

We would recommend using LaTeX for writing your submission but also accept handwritten solutions, but please note that if we can not read or understand it, we cannot grade it.

To get full points, always provide the steps in your derivation/proofs and make clear when/how you use known results, for example, from the lecture (e.g. already proven concentration inequalities).

Exercise 1.1: Bayes Risk I

Let $\mathcal{X} = \mathcal{Y} = \{1, 2, 3\}$. Assume the labels Y follow the distribution

$$P(Y = j) = \begin{cases} 1/4 & \text{if } j = 1, 2 \\ 1/2 & \text{if } j = 3 \end{cases}$$

Conditioned on the labels, the features X are distributed as

$$P(X = i|Y = 1) = \begin{cases} 1/3 & \text{if } i = 2 \\ 2/3 & \text{if } i = 3 \end{cases}$$

$$P(X = i|Y = 2) = \begin{cases} 1/2 & \text{if } i = 1 \\ 1/2 & \text{if } i = 3 \end{cases}$$

$$P(X = i|Y = 3) = \begin{cases} 2/3 & \text{if } i = 1 \\ 1/3 & \text{if } i = 2 \end{cases}$$

(a) Compute the Bayes classifier (i.e. the classifier h^* that maximizes the probability that $h^*(x) = y$ for any given x).

(b) Compute the Bayes risk.

Hint: You don't have to compute the marginals of X.

(3+2 = 5 points)

Exercise 1.2: Bayes Risk II

Assume $\alpha \leq \frac{1}{3}$. Let $\mathcal{X} = \{1, 2, \dots, 30\}$, $\mathcal{Y} = \{\pm 1\}$ and class probability be

$$\eta(x) = \mathbb{P}(y = +1|x) = \begin{cases} 1 - \alpha & \text{if } x \in \{11, 12, \dots, 20\} \\ \alpha & \text{otherwise.} \end{cases}$$

You may assume that x is sampled from a probability mass function $\mathbf{p} = (p_1, p_2, \dots, p_{30})$.

a) Compute the Bayes risk for the problem, and show that the Bayes classifier is in the class of signed intervals

$$\mathcal{H}_{int} = \left\{ h_{s,t,b}(x) = b \text{ for } x \in (s, t), \text{ and } -b \text{ otherwise } : b \in \{\pm 1\}, s, t \in \mathbb{R}, s < t \right\}$$

b) Find the minimal risk achieved by the class of decision stumps

$$\mathcal{H}_{ds} = \left\{ h_{t,b}(x) = b \text{ for } x < t, \text{ and } -b \text{ otherwise } : b \in \{\pm 1\}, t \in \mathbb{R} \right\}.$$

The solution will depend on $\mathbf{p}$. Give possible optimal decision stumps.

Hint: It may help to define $q_1 = (p_1 + \ldots + p_{10})$, $q_2 = (p_{11} + \ldots + p_{20})$, $q_3 = (p_{21} + \ldots + p_{30})$.

(2 + 3 = 5 points)

Exercise 1.3: Universal consistency of ϵ-neighbourhood classifiers

Assume that the domain $\mathcal{X} \subseteq \mathbb{R}$. Given a training sample $S = \{(x_1, y_1), \ldots, (x_m, y_m)\} \subset \mathcal{X} \times \{\pm 1\}$ and $\epsilon > 0$, an ϵ-neighbourhood classifier is defined as $h_{S,\epsilon} : \mathcal{X} \to \{\pm 1\}$ such that $h_{S,\epsilon}(x) = \operatorname{sign}\left(\sum_{i\,:\,|x_i - x| \leq \epsilon} y_i\right)$.

You may assume that $\operatorname{sign}(0) = -1$ and the summation is zero if there is no x_i that is ϵ-close to x.

(a) For any fixed $\epsilon > 0$ and training sample S, express $h_{S,\epsilon}$ as a plug-in classifier with a weighted average estimator $\widehat{\eta}$.

(b) In the next two subproblems, we prove Universal consistency in a specific setting for $\mathcal{X} = \{0, 1\}$ and $\epsilon < 1$.

Simplify the weighted average estimator $\widehat{\eta}$ in this case, and show that, for every $x \in \{0, 1\}$, $\widehat{\eta}(x)$ converges to $\eta(x)$ in probability as $m \to \infty$.

Hint: If $Z \sim \text{Binomial}(n, p)$ then $Z/n \to p$ in probability as $n \to \infty$.

(c) Use part (b) to show that ϵ-neighbourhood classifier is universally consistent on $\mathcal{X} = \{0, 1\}$ for any $\epsilon < 1$ without using Stone's theorem.

Hint: Use the risk bound for plug-in classifiers from the lecture.

(2+2+3 = 7 points)