Pattern Recognition: Homework 5

Due date: 2023.3.28

Problem 1 (20 pt)

In this problem, you need to deduce a closed-form Bayesian classifier for a synthetic distribution. Suppose we have two class ω_1, ω_2, and the prior probability is $P(\omega_1) = \frac{2}{3}, P(\omega_2) = \frac{1}{3}$. We can observe a two-dimensional feature $(x, y) \in \mathbb{R}^2$ for each data. The class-conditional feature density follows the Gaussian distribution below

$$P(x, y|\omega = \omega_1) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

$$P(x, y|\omega = \omega_2) = \frac{1}{2\pi} e^{-\frac{(x-2)^2+(y-2)^2}{2}}$$

Please give a Bayesian optimal classifier $f(\cdot) : \mathbb{R}^2 \mapsto \{\omega_1, \omega_2\}$ that predicts the class from the observable feature (x, y), and draw its decision boundary on the $\mathbb{R}^2$ plane.

Problem 2 (20 pt)

In the class, we introduce non-uniform loss, where the cost of making different mistakes differs. Now let's formalize it. Suppose the cost(loss) for making a decision $\omega = \omega_2$ while the real type is ω_1 equals λ_1, and the cost for deciding $\omega = \omega_1$ when the real type is ω_2 equals λ_2. Recall that the goal of the Bayesian decision is to achieve minimum expected cost, which is

$$f^* = \min_f \int_{\boldsymbol{x}} \sum_{i=1}^{2} \lambda_i \cdot \mathbb{I}(f(x) \neq \omega_i) \cdot P(\omega_i, \boldsymbol{x}) \mathrm{d}x.$$

Please prove that the final classifier outputs $f^*(\boldsymbol{x}) = \omega_1$ if and only if $\frac{P(\omega_1|\boldsymbol{x})}{P(\omega_2|\boldsymbol{x})} \geq \frac{\lambda_2}{\lambda_1}$. Specially, when $\lambda_1 = \lambda_2$, this corresponds to our original max-posterior-solution. (Don't be scared, it is an easy problem.)

Problem 3 (20 pt)

Suppose the hypothesis class is $\mathcal{H} = \left\{h(x) = (-1)^{k+\mathbb{I}[a \leq x \leq b]} : k \in \{0, 1\}, a, b \in \mathbb{R}, a \leq b\right\}$. Intuitively, you can understand it as an "interval predictor", which selects an interval $[a, b]$ as one class and predicts the remaining real numbers $(-\infty, a) \cup (b, \infty)$ as another class. Prove that the VC dimension of $\mathcal{H}$ is 3.

Problem 4, Programming (40 pt)

Generate 1000 samples from the distribution of $\boldsymbol{x}$ in problem 1. You can first sample a random variable that has 2/3 probability to be 0, and 1/3 probability to be 1. If it is 0, then generate $\boldsymbol{x} = (x, y)$ from $P(x, y|\omega = \omega_1)$, otherwise from distribution $P(x, y|\omega = \omega_2)$.

- Use Parzen window of proper width to estimate the density function of $P(x, y)$. Plot the two-dimensional density as heatmap. (hint: you can only care about a bounded region like $[-5, 5] \times [-5, 5]$ and discretize it into grids of width 0.02. Finally, draw that matrix as a heatmap.)
- Change the window width and kernel function, plot the heatmap, and calculate the mean density error $\int_x \int_y |f(x, y) - P(x, y)| \, \mathrm{d}x \, \mathrm{d}y$, where $f(x, y)$ is your predicted density at certain grid $(x, x + \delta x) \times (y, y + \delta y)$ and $P(x, y)$ is real density. What are the best compositions?