DSC 140A - Homework 06

Due: Wednesday, February 21

Problem 1.

Let X be a continuous random variable, and let Y be a binary random variable. Suppose the class conditional densities are know to be:

$$p_X(x \mid Y = 0) = \begin{cases} 1/5, & \text{if } 0 \leq x \leq 2 \\ 1/3, & \text{if } 2 < x \leq 3 \\ 1/15, & \text{if } 3 < x \leq 7 \end{cases}$$

$$p_X(x \mid Y = 1) = \begin{cases} 1/6, & \text{if } 0 \leq x \leq 1 \\ 1/8, & \text{if } 1 < x \leq 5 \\ 1/6, & \text{if } 5 < x \leq 7 \end{cases}$$

Suppose also that $\mathbb{P}(Y = 0) = 0.4$ and $\mathbb{P}(Y = 1) = 0.6$.

For what values of $x \in [0, 7]$ will the Bayes classifier predict $y = 1$?

Problem 2.

Let X be a continuous random variable, and let Y be a random class label (1 or 0). Recall that the Gaussian probability density function (pdf) is given by:

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Suppose $p_X(x \mid Y = 1)$ is a Gaussian pdf with $\mu = 2$ and $\sigma = 3$ and that $p_X(x \mid Y = 0)$ is also Gaussian with $\mu = 5$ and $\sigma = 3$. Suppose, too, that $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}$.

Recall that the Bayes error is the probability that the Bayes classifier makes an incorrect prediction. What is the Bayes error for this distribution? Show your work.

Hint: you'll want some way to compute the area under a Gaussian. You can use the tables that appear in the back of your statistics book, or you can use something like `scipy.stats.norm.cdf`. We'll let you read the documentation to see how to use it, but it may be helpful to remember that if F is the cumulative density function for a distribution with density f, then $\int_a^b f(x)\,dx = F(b) - F(a)$.

Problem 3.

The file linked below contains a data set of 150 samples. The first column contains a single continuous feature, X, assumed to have been drawn from an unknown probability density. The second column contains the binary class label Y.

`https://f000.backblazeb2.com/file/jeldridge-data/011-univariate_density_estimation/data.csv`

In this problem, use a histogram estimator with bins $[0, 1.5), [1.5, 3), \ldots, [13.5, 15)$ to estimate the requested probabilities. In each part, show your code and provide your reasoning.

Hint: you may find `np.histogram` useful.

a) Estimate $\mathbb{P}(Y = 1 \mid X = 6.271)$ directly.

b) Estimate $\mathbb{P}(Y = 1 \,|\, X = 6.271)$ by estimating all of: 1) the marginal density $p_X(x)$, 2) the class-conditional density $p_X(x \,|\, Y = 1)$, and 3) the class prior $\mathbb{P}(Y = 1)$, and then applying Bayes' rule.

c) For what values of $x \in [0, 15]$ will the Bayes classifier predict $y = 1$?