DSC 140A - Discussion 04

Problem 1.

Suppose the following data are observed:

x	y
2.1	1
4.7	1
2.3	0
0.8	0
1.3	1
5.2	1
7.4	0
9.4	1
3.9	1

You may assume that the x-values were drawn from a continuous distribution, and the y-values represent the label of each point.

To estimate all probabilities below, use a histogram estimator with 5 equally-sized bins spanning the interval from 0 to 10. The bins should include their starting point and exclude their ending point.

a) What is the estimated density $p_X(x)$ at $x = 3$?

b) What is the estimated probability that a new point x is in the interval $[3, 4]$?

c) What is the estimated conditional density $p(x \mid Y = 1)$ at the point $x = 2$?

d) Using the Bayes classification rule, what is the predicted label y of a new point $x = 2.5$?

Problem 2.

The Rayleigh distribution has pdf:

$$p(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)},$$

where σ is a parameter.

Suppose a data set of independent points $x_1, \ldots, x_n$ is drawn from a Rayleigh distribution with unknown parameter σ. Show that the log-likelihood of σ given this data is:

$$L(\sigma | x_1, \ldots, x_n) = n \log \frac{1}{\sigma^2} + \sum_{i=1}^{n} \log x_i - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2$$