# DSC 140A - Homework 07

Due: Wednesday, March 1

**Problem 1.**

The Rayleigh distribution has pdf:

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

where $\sigma$ is a parameter.

Suppose a data set of points $x_1, \ldots, x_n$ is drawn from a Rayleigh distribution with unknown parameter $\sigma$. It was shown in discussion section that the log-likelihood of $\sigma$ given this data is:

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

Show that the maximum likelihood estimate of $\sigma$ is:

$$\sigma_{\text{MLE}} = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} x_i^2}.$$

**Problem 2.**

Let $X$ be a continuous random variable, and let $Y$ be a random class label (1 or 0).

Suppose $p_X(x \mid Y = 1)$ is Gaussian 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 at the link below contains a data set of 100 points from two classes (1 and -1).

`https://f000.backblazeb2.com/file/jeldridge-data/003-two_clusters/data.csv`

The first two columns contains features, and the last column contains the label of the point. Note that the labels are 1 and -1, not 1 and 0, and that there are no column headers.

In all parts of this problem you may use code to compute your answers. If you do, be sure to include your code.

**a)** Suppose two Gaussians with full covariance matrices are used to model the densities $p_X(x \mid Y = 1)$ and $p_X(x \mid Y = -1)$. What are the maximum likelihood estimates for the covariance matrices of each Gaussian?

(Allow each Gaussian to have its own covariance matrix; don't use the same covariance for both.)

*Hint*: the covariance matrix for the Gaussian fit to points from class 1 should have 12.29 in its top-left entry.

**b)** Using the estimated Gaussians with the Bayes classification rule, what are the predicted labels of each of the following points?

- $(0, 0)^T$
- $(1, 1)^T$
- $(10, 5)^T$
- $(5, -5)^T$
- $(8, 5)^T$

Show your calculations.

*Note*: making predictions in this way (using Gaussians with unequal covariance matrices) is known as *Quadratic Discriminant Analysis*.