Spring 2020 318 711 Homework 2

Use programming tool if needed.

Problem 1

Suppose there are two types of fishes, "big" and "small." The length of a small fish is random and uniformly distributed in $[0, 5]$ and the length of a big fish is also random and is uniformly distributed in $[2, 7]$. Suppose the population ratio of the small fish to big fish is 1 to 1.

a.) Find the optimal fish classifier, its decision regions and boundary, and the probability of error.

b.) How does your answer change if the population ratio (small to big) is changed to 2 to 1?

Problem 2

In a two-class classification problem, the two classes are equally likely and $p(x|H_k)$, $k = 0, 1$, are two-dimensional Gaussians, with m_k, Σ_k, with $m_k = [2k, 2k], \Sigma_k = \sigma^2 I$ (I is a 2×2 identity matrix).

a.) Find the optimal classifier and its decision regions and boundary. Plot or sketch the decision boundary.

b.) Find the probability of error. You may use the Q function (tail function).

c.) Find the probability of error for $\sigma^2 = 1/9$ and $\sigma^2 = 1/16$. What conclusions can you draw from this?

Problem 3

Repeat Problem 3 when the two Gaussians are given by $m_0 = [0, 0]^T, m_1 = [1, 0]^T$ and $\Sigma_0 = [0.1, 0; 0, 0.15], \Sigma_1 = 0.2, 0; 0, 0.25]$. Unlike Problem 3, the probability of error might be difficult to find analytically for this problem. If this is the case, can you find a way to estimate the probability of error? If you can, show your result. If not, why not?

Problem 4

For a two-class classification problem, suppose the two populations (H_0 and H_1) are given by $N(-2, 4)$ and $N(2, 4)$.

a.) Find the log-likelihood ratio test and plot the ROC curve.

b.) How does the ROC curve change if the variance of the Gaussian increases (e.g., from 1 to 2)? Similarly what happens if the variance decreases?

Problem 5

Let $f(x) = x^2 - 7x + 6$ and consider the following gradient descent algorithm

$$x_{n+1} = x_n - \epsilon f'(x_n) \tag{1}$$

where $\epsilon < 1$ is a small positive number and $f'(x)$ is the derivative of $f(x)$.

a.) What is x_n as $n \to +\infty$? Provide a derivation or proof.

b.) Simulate the gradient algorithm, does it agree your results in part a)?

Problem 6

Apply logistic regression to the data set in folder "Problem 5" (train on the training data and test the resulting logistic regression model on the testing set). Show your w and classification error rate on both the training and testing set. How do you know if your result is reasonable?

Problem 7

Repeat Problem 6 on the data set of "Problem 6". Does logistic regression work well in this case and why?