CSE 455/555 Spring 2013 Homework 2: Bayesian Decision Theory

Jason J. Corso
Computer Science and Engineering
SUNY at Buffalo
jcorso@buffalo.edu
Solution provided by TA Yingbo Zhu

This assignment does not need to be submitted and will not be graded, but students are advised to work through the problems to ensure they understand the material.

You are both allowed and encouraged to work in groups on this and other homework assignments in this class.

Programming Problem

Consider the two-dimensional datapoints from two classes ω_1 and ω_2 below, and each of them come from a Gaussian distribution $p(x|\omega_k) \sim \mathcal{N}(\mu_k, \Sigma_k)$.

Table 1: Data points from class ω_1 and ω_2

ω_1	ω_2
$(0,0)$	$(6,9)$
$(0,1)$	$(8,9)$
$(2,2)$	$(9,8)$
$(3,1)$	$(9,9)$
$(3,2)$	$(9,10)$
$(3,3)$	$(8,11)$

1. What is the prior probability for each class, *i.e.* $p(\omega_1)$ and $p(\omega_2)$.

$$p(\omega_1) = \frac{6}{12} = 0.5 \qquad p(\omega_2) = \frac{6}{12} = 0.5$$

2. Calculate the mean and covariance matrix for each class.
   the following matlab code will calculate the corresponding value.
   class1 = [0, 0; 0, 1; 2, 2; 3, 1; 3, 2; 3, 3;];
   class2 = [6, 9; 8, 9; 9, 8; 9, 9; 9, 10; 8, 11;];
   mu1 = mean(class1);
   mu2 = mean(class2);
   sigma1 = cov(class1);
   sigma2 = cov(class2);

$$\mu_1 = (1.8333, 1.5) \qquad \mu_2 = (8.1667, 9.3333)$$

$$\Sigma_1 = \begin{bmatrix} 2.1667 & 1.1000 \\ 1.1000 & 1.1000 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1.3667 & -0.0667 \\ -0.0667 & 1.0667 \end{bmatrix}$$

The above covariance matrix is the unbiased estimate (*i.e.* normalized with $1/(N-1)$).

3. Derive the equation for the decision boundary that separates these two classes, and plot the boundary. (Hint: you may want to use the posterior probability $p(\omega_1|x)$)

$$p(\omega_i|x) \propto p(x|\omega_i)p(\omega_i)$$

For minimum error rate classification, we can write the discriminant function for each class as

$$g_i(x) = p(x|\omega_i)p(\omega_i)$$

since this question involves Gaussian distribution, apply natural logrithm will make it easier for us, i.e. we can write the discriminant function as

$$g_i(x) = \ln p(x|\omega_i) + \ln p(\omega_i)$$

so the discriminant function in this case would be

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln p(\omega_i)$$

The decision boundary would be the solution of $g_1(x) = g_2(x)$.
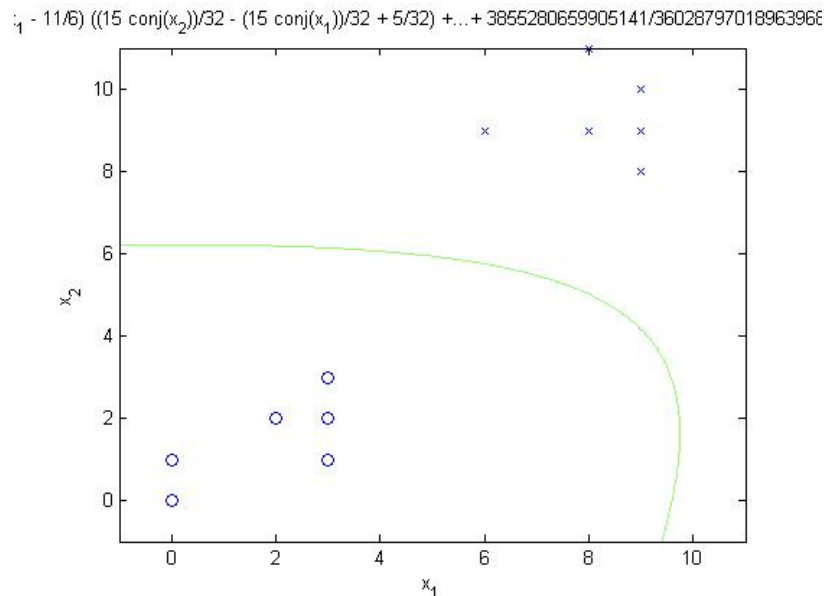
The following code implements the above
syms x1 x2;
g1 = -0.5*([x1;x2]-mu1')'*inv(sigma1)*([x1;x2]-mu1')-log(2*pi)-0.5*log(det(sigma1));
g2 = -0.5*([x1;x2]-mu2')'*inv(sigma2)*([x1;x2]-mu2')-log(2*pi)-0.5*log(det(sigma2));

g = g1 - g2;
ezplot(g, [[-1,11], [-1,11]]);
Term $\ln p(\omega_i)$ is omitted since we have equal prior in this problem. The ploted boundary looks like the following

4. Think of the case that the penalties for misclassification are different for the two classes (*i.e.* not zero-one loss), will it affect the decision boundary, and how?
   Let's consider the case that misclassify a sample to class 2 cost more than misclassify a sample to class 1. Then the decision boundary will move toward class 2, i.e. more samples will be classified as class 1. Since we can afford more errors in the case that we misclassify samples to class 1.

Mathematical Problem

Consider two classes ω_1 and ω_2 in pattern space Ω with continuous probability distribution $p_1(x)$ and $p_2(x)$, respectively. This two-class classification problem can be interpreted as dividing space Ω into two exhaustive and disjoint sets Ω_1 and Ω_2, such that $\Omega_1 \cup \Omega_2 = \Omega$ and $\Omega_1 \cap \Omega_2 = \emptyset$. If $x_i \in \Omega_k$ then assign x_i to class ω_k.

1. Suppose you are given a discriminant function $f(\cdot)$, list the two errors this function can make.

$$f(x) = \begin{cases} \omega_2, \text{when} x \in \omega_1 \\ \omega_1, \text{when} x \in \omega_2 \end{cases}$$

2. Write down the probability of error corresponds to the two erros.

$$\int_{\Omega_2} p_1(x)dx \qquad \text{and} \qquad \int_{\Omega_1} p_2(x)dx$$

3. Suppose the costs for the two types of errors are c_1 and c_2, write down the total expected cost.
   Assume that c_1 is the cost of the first error and c_2 is the cost of the second error. Then the expected cost would be

$$c_1 p(\omega_1) \int_{\Omega_2} p_1(x)dx + c_2 \left(1 - p(\omega_1)\right) \int_{\Omega_1} p_2(x)dx$$