Solutions to Homework Set Three
ECE 175
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos

1.a) The constants are found by setting the integral of the PDF equal to 1.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} K_j e^{-\frac{|x-a_j|}{b_j}} dx \\ &= 2K_j \int_0^{\infty} e^{-\frac{x}{b_j}} dx \\ &= 2K_j b_j \end{aligned}$$

So $K_j = 1/(2b_j)$.

b) Bayes decision rule for this binary classification gives us

$$P_{X|Y}(x|0)P_Y(0) \underset{y=1}{\overset{y=0}{\gtrless}} P_{X|Y}(x|1)P_Y(1)$$

$$\frac{P_{X|Y}(x|0)}{P_{X|Y}(x|1)} \underset{y=1}{\overset{y=0}{\gtrless}} \frac{P_Y(1)}{P_Y(0)}$$

$$\frac{K_0}{K_1}\exp\left(-\frac{|x-a_0|}{b_0} + \frac{|x-a_1|}{b_1}\right) \underset{y=1}{\overset{y=0}{\gtrless}} \frac{P_Y(1)}{P_Y(0)}$$

$$-\frac{|x-a_0|}{b_0} + \frac{|x-a_1|}{b_1} \underset{y=1}{\overset{y=0}{\gtrless}} \ln\left(\frac{P_Y(1)K_1}{P_Y(0)K_0}\right)$$

$$\frac{|x-a_0|}{b_0} - \frac{|x-a_1|}{b_1} \underset{y=0}{\overset{y=1}{\gtrless}} \ln\left(\frac{b_1 P_Y(0)}{b_0 P_Y(1)}\right)$$

c) For $a_0 = 0$, $a_1 = 1$, $b_0 = 1$, $b_1 = 2$, and $P_Y(0) = P_Y(1) = 0.5$ we can find

$$|x| - \frac{|x-1|}{2} \underset{y=0}{\overset{y=1}{\gtrless}} \ln(2)$$

$$|x| - \frac{|x-1|}{2} - \ln(2) \underset{y=0}{\overset{y=1}{\gtrless}} 0.$$

A plot of the decision function is shown in Figure 1. We choose $y = 0$ for $-2.386 \le x \le 0.795$.

d) Changing $P_Y(0) = 0.75$ and $P_Y(1) = 0.25$ we get

$$|x| - \frac{|x-1|}{2} - \ln(6) \underset{y=0}{\overset{y=1}{\gtrless}} 0.$$

A plot of the decision function is also shown in Figure 1. We choose $y = 0$ for $-4.584 \le x \le 2.584$.

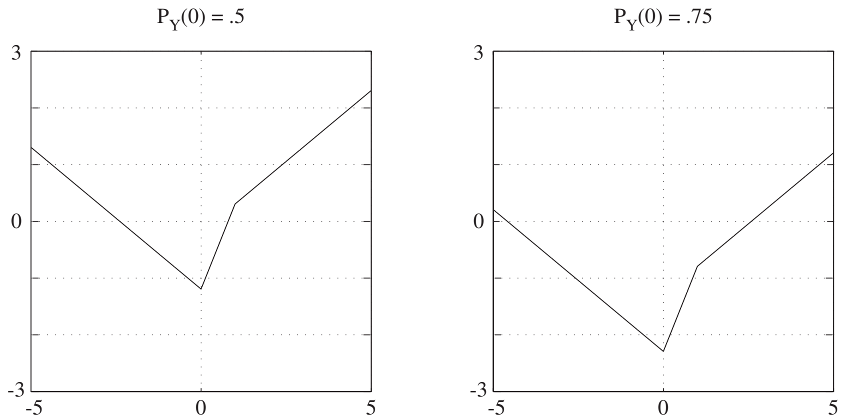


Figure 1: Decision functions for Problems 1c) and 1d). Negative values indicate selection of $y = 0$.

2.a) Setting the conditional densities equal to solve for the decision boundary give us

$$\begin{aligned}
P_Y(0)\exp(-\frac{1}{2\sigma^2}(\mathbf{x}-\mu_0)^T(\mathbf{x}-\mu_0)) &= P_Y(1)\exp(-\frac{1}{2\sigma^2}(\mathbf{x}-\mu_1)^T(\mathbf{x}-\mu_1)) \\
\frac{-\|\mathbf{x}-\mu_0\|^2}{2\sigma^2} + \ln(P_Y(0)) &= \frac{-\|\mathbf{x}-\mu_1\|^2}{2\sigma^2} + \ln(P_Y(1)) \\
\frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x} - 2\mu_0^T\mathbf{x} + \mu_0^T\mu_0) - \ln(P_Y(0)) &= \frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x} - 2\mu_1^T\mathbf{x} + \mu_1^T\mu_1) - \ln(P_Y(1)) \\
(\mu_1-\mu_0)^T\mathbf{x} - \left(\frac{\mu_1^T\mu_1 - \mu_0^T\mu_0}{2} + \sigma^2\ln\left(\frac{P_Y(0)}{P_Y(1)}\right)\right) &= 0 \\
(\mu_1-\mu_0)^T\left(\mathbf{x} - \left(\frac{\mu_1+\mu_0}{2} + \frac{\sigma^2}{\|\mu_1-\mu_0\|^2}\ln\left(\frac{P_Y(0)}{P_Y(1)}\right)(\mu_1-\mu_0)\right)\right) &= 0.
\end{aligned}$$

This is the form we want. We see that

$$\mathbf{w} = \mu_1 - \mu_0$$

and

$$\mathbf{x}_0 = \frac{\mu_1+\mu_0}{2} + \frac{\sigma^2}{\|\mu_1-\mu_0\|^2}\ln\left(\frac{P_Y(0)}{P_Y(1)}\right)(\mu_1-\mu_0).$$

b) Figure 2 shows the contours and decision boundaries.

c) The vector $\mathbf{w}$ is orthogonal to the decision boundary (a line in this 2D case). When the covariances are the identity, this vector lies on the line that joins the two mean vectors. The point $\mathbf{x}_0$ is the point on the decision boundary that intersects the line connecting μ_0 and μ_1. When $P_Y(0) = P_Y(1) = 0.5$, the hyperplane bisects the line segment between the two distribution centers. Changing the priors moves the intersection point along the line that joins the means. This makes it more difficult to pick one class and, of course, easier to pick the other. This result extends to the N dimensional case as well.

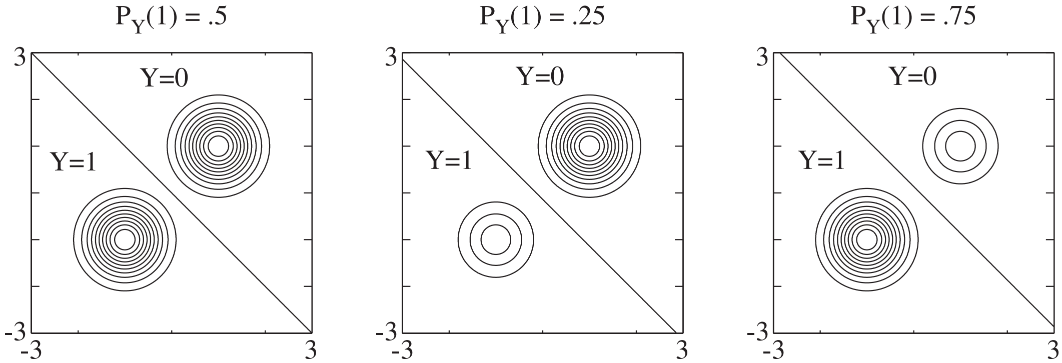


Figure 2: Conditional probability contours with decision region boundary.

3)

1. Means for every class.

Figure 3: Mean images, as averaged from the training dataset for each digit

2. Gaussian Classification using BDR

   (a) Error rates for each class, i.e. P(Error|Class = i) for all $i = 0 \dots 9$.

Table 1: Probability of error for each class

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.14	**0.02**	**0.34**	**0.20**	**0.21**	**0.34**	**0.23**	**0.24**	**0.25**	**0.26**

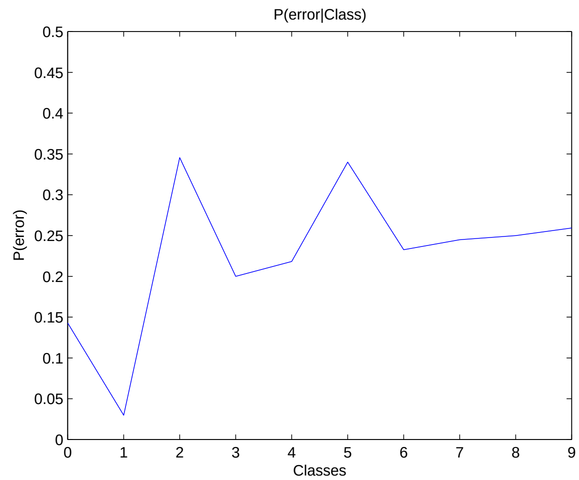
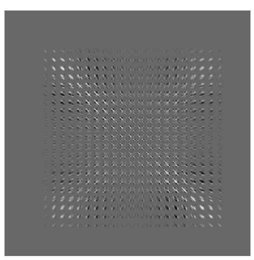


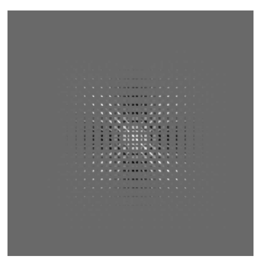
Figure 4: Probability of error for each class

   (b) Total error rate, P(error) = 0.222

3. (Optional exercise)



Covaricance for digit0



Covaricance for digit1

We see that most of the entries of the covariance matrix is zero (gray values are zero), which makes it a singular matrix, and hence the inverse does not exist. Later you will learn techniques how to tackle with this problem using the method of Principal Component Analysis.