Department of Computer Science
California State University, Channel Islands

MATHCOMPPH-546: Lesson 5 phys546 Classification HW_5A

Student Name: Sandipta Subir Khare
Student Major: Computer Science

5.1 Suppose that the class-conditional probability functions for o1 and o2 are Gaussians with (mi, si) of (4, 2) and (10, 1), and that they have equal prior probabilities (P1 ¼ P2 ¼ ½). What is the optimal decision threshold? (Try by calculation, and then check using CondProb.xls).

Ans.

Page No.
Date

5.1 The optimal decision rule classifies an instance to class, which has the highest posterior, the optimal threshold is just point, where the posteriors match.

$\Rightarrow$ The optimal threshold satisfy

$$P(c/\omega_1)\,P_1 = P(c/\omega_2)\,P_2$$

Given, $P_1 = P_2 = \frac{1}{2} = 0.5$

Thus, $P(c/\omega_1) = P(c/\omega_2)$

$$\Rightarrow \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{1}{2 \times 2^2}(c-4)^2\right) =$$

$$\frac{1}{1\sqrt{2\pi}} \exp\left(-\frac{1}{2 \times 1^2}(c-10)^2\right)$$

Since ω_1 and ω_2 are Gaussians with (μ_i, σ_i) of $(4, 2)$ and $(10, 1)$

$$\Rightarrow \frac{(c-4)^2}{8} = -\log 2 + \frac{(c-10)^2}{2}$$

$$\Rightarrow (c-4)^2 = -8\log 2 + 2(c-10)^2$$

$$\Rightarrow c^2 - 8c + 16 = -5.54 + 4c^2 - 80c + 400$$

$$\Rightarrow 3c^2 - 72c + 378.46 = 0$$

$$\Rightarrow c = 7.775,\ 16.224$$

Page No.
Date

Since, $x \in (7.775, 16.224) \Leftrightarrow P(c|\omega_2) > P(c|\omega_1)$

$\Rightarrow$ Therefore, the optimal rule is to classify x to ω_2,
if $x \in (7.775, 16.224)$, or otherwise, to ω_1.

5.2 What are the decision thresholds for two class-conditional probabilities which are Gaussian in shape, with means m1 ¼ 4 and m2 ¼ 10, variances s2 1 ¼ 4 and s2 2 ¼ 1, and prior probabilities P(o1) ¼ 2/3 and P(o2) ¼ 1/3? (Try by calculation, and then check using CondProb.xls).

Ans.

Prior Probabilities:

$$P(\omega_1) = \frac{2}{3} \qquad P(\omega_2) = \frac{1}{3}$$

$$P(X|\omega_1) = Normal(4,4) \qquad P(X|\omega_2) = Normal(10,1)$$

By Bayes' theorem

$P(\omega_1|X) = \frac{P(X|\omega_1)*P(\omega_1)}{P(X)}$

The decision boundary is x^* such that

$P(\omega_1|X = x^*) = P(\omega_2|X = x^*)$

$\Rightarrow \frac{(P(x^*|\omega_1) \times P(\omega_1))}{P(X = x^*)} = \frac{(P(x^*|\omega_2) \times P(\omega_2))}{P(X = x^*)}$

$\Rightarrow \left(P(x^*|\omega_1) \times \frac{2}{3}\right) = \left(P(x^*|\omega_1) \times \frac{1}{3}\right)$

$\Rightarrow \frac{2}{3}\frac{1}{\sqrt{2\pi 4}} e^{-\frac{1}{2*4}(x^*-4)^2} = \frac{1}{3}\frac{1}{\sqrt{2\pi 1}} e^{-\frac{1}{2*1}(x^*-10)^2}$

$\Rightarrow e^{-\frac{1}{2*4}(x^*-4)^2} = e^{-\frac{1}{2*1}(x^*-10)^2}$

$\Rightarrow \frac{1}{2*4}(x^* - 4)^2 = \frac{1}{2*1}(x^* - 10)^2$

$\Rightarrow \frac{1}{4}((x^*)^2 + 16 - 8x^*) = ((x^*)^2 + 100 - 20x^*)$

$\Rightarrow ((x^*)^2 + 16 - 8x^*) = (4(x^*)^2 + 400 - 80x^*)$

$\Rightarrow 3(x^*)^2 + 384 - 72x^* = 0$

$\Rightarrow 3(x^*)^2 + 384 - 48x^* - 24x^* = 0$

$\Rightarrow 3(x^*)^2 - 24x^* + 384 - 48x^* = 0$

$\Rightarrow 3x^*(x^* - 8) - 48(x^* - 8) = 0$

$\Rightarrow 3(x^* - 16)(x^* - 8) = 0$

Thus boundaries are $x^* = 16$ and $x^* = 8$

5.3 Select the optimal decision where the class-conditional probabilities are Gaussians (i.e., N(m, s2)), given by N(2, 0.5) and N(1.5, 0.2) and the corresponding priors are 2/3 and 1/3. (Try by calculation, and then check using CondProb.xls).

Ans.

Given that the class-conditional probabilities are Gaussian given by N(2 0.5) and N(15, 0.2) and the corresponding priors are 2/3 and 13. We have to select the optimal decision.

Let Select the optimal decision where:

$\Omega = \{\omega_1, \omega_2\}$

$P(x|\omega_1)$ follows N(2, 0.5)

$P(x|\omega_2)$ follows N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

$\lambda = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

Using Likelihood ratio Decision Rule:

Decide ω1 if:

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

and decide ω2 otherwise

i.e., If

$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Then decide ω_1 Otherwise decide ω_2

here$\lambda_{11=1}$ λ_{12} =2

$\lambda_{21=3}$ $\lambda_{22=4}$

Now putting values in R.H.S ,we get

$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > -1/2$$

And as P(x | ω1) follows N(2, 0.5) and P(x | ω2) follows N(1.5, 0.2),

Graph of N(2, 0.5)>graph of N(1.5, 0.2)

That's why $\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)}$ will always greater than 1.

So ,we decide ω_1 .

ω_1 is the optimal decision.

5.4 For the fish canning problem discussed in this chapter, customers may not mind some of the more expensive salmon turning up in their cans of sea bass, but may have more objections to misclassified sea bass turning up in their salmon cans. We can factor this into the analysis with the loss function: which term should we set to be the larger ... l12 [the loss associated with misclassifying a sea bass (o1) as salmon (o2)] or l21 [the loss associated with misclassifying a salmon (o2) as a sea bass (o1)]?

Ans.

The Mahalanobis distance of an obserevation from the population with mean μ and variance-covariance matrix Σ is

$$(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)'.$$

Thus, with x = [1.0,2.2] and $\mu_1 = [0,0]$ and $\mu_2 = [3,3]$ is carried out in the following R program:

```
> DS <- matrix(c(1.1,.3,.3,1.9),nrow=2)
> DS
    [,1] [,2]
[1,] 1.1 0.3
[2,] 0.3 1.9
> mu1 <- c(0,0); mu2 <- c(3,3)
> x <- c(1,2.2)
> (x-mu1)%*%solve(DS)%*%t(x-mu1)
Error in (x - mu1) %*% solve(DS) %*% t(x - mu1) :
non-conformable arguments
> t(x-mu1)%*%solve(DS)%*%(x-mu1)
     [,1]
[1,] 2.952
> t(x-mu2)%*%solve(DS)%*%(x-mu2)
     [,1]
[1,] 3.672
```

Thus, the vector is classified to the population with mean $\mu_1 = [0,0]$.

5.6 In a two-class, two-feature classification task, the feature vectors described by Gaussian distributions with the same covariance matrix

$$\Sigma = \begin{vmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{vmatrix}$$

And the means are m1 [0, 0] and m2 ¼ [3, 3]. Classify the vector [1.0, 2.2] according to the Bayesian classifier. (Hint: use the Mahalanobis distances to the two means).

Ans.

I have done it in R.

```
> DS <- matrix(c(.4,.3,.3,.4),nrow=2)
> DS
     [,1] [,2]
[1,]  0.4  0.3
[2,]  0.3  0.4
> mu1 <- c(0,0); mu2 <- c(3,3)
> x <- c(1,2.2)
> t(x-mu1)%*%solve(DS)%*%(x-mu1)
         [,1]
[1,] 14.51429
> t(x-mu2)%*%solve(DS)%*%(x-mu2)
     [,1]
[1,] 12.8
```

- So, vector is classified to the population with mean µ1=[3, 3]