Ryerson University
Department of Electrical
& Computer Engineering
Faculty of Engineering & Architectural Science

Course Title:	Intelligent Systems
Course Number:	ELE888
Semester/Year (e.g.F2016)	W2020

Instructor:	Ling Guan

Assignment/Lab Number:	1
Assignment/Lab Title:	Bayesian Decision Theory

Submission Date:	10/02/2020
Due Date:	10/02/2020

Student LAST Name	Student FIRST Name	Student Number	Section	Signature*
Kim	Sasung	500642700	3	[signature]
Chen	Jie	500715242	3	[signature]
Dominic	Paul	500636052	3	P. D.

Reset Form

*By signing above you attest that you have contributed to this written lab report and confirm that all work you have contributed to this lab report is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a "0" on the work, an "F" in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at: http://www.ryerson.ca/senate/current/pol60.pdf

1. Using a single discriminant function g(x2), design a 2-class minimum-error-rate classifier (dichotomizer) from the given data, to classify IRIS samples into either Iris Setosa or Iris Versicolour, according to the feature: *sepal width.*
2. Using the shell programlab1.m, write a program that will take an individual sample value as the input and will return the posterior probabilities and the value of g(x2)
3. Identify the class labels for the feature values using your program, and indicate their respective posterior probabilities and discriminant function values: x1= [3.3, 4.4, 5.0, 5.7, 6.3]

Since the range of sepal width for w1 is between 3 to 4.4 and w2 is 2 to 3.5, and the lowest number in the given dataset is 3.3, the probability of choosing w2 is much lower than probability of choosing w1 for every testing set. Thus, the result is always Iris-Setosa (w1) with Sepal Width.

Table 1: Classifying samples by label, posterior, and discriminate function according to sepal width

X_2	Class	**Posterior Probabilities [P(w_1|x),P(w_2|x)]**	**Discriminate Function [g(x)]**
3.3	w1	[0.8281, 0.1719]	0.6563
4.4	w1	[1.0000, 0.0000]	0.9999
5.0	w1	[1.0000, 0.0000]	1.0000
5.7	w1	[1.0000, 0.0000]	1.0000
6.3	w1	[1.0000, 0.0000]	1.0000

4. Arrive at an optimal threshold (Th1) that separates classes w1 and w2 (theoretically or experimentally). Justify your result.
5. Suggest how T_{h1} would be a effected if a higher penalty is associated with classifying class w2 as class w1 – show with experiment

The effect of a higher penalty associated with classifying class w2 as w2 requires the Bayes Risk Formula which uses the loss function to describe the penalties for misclassification. We are assuming that the penalty associated with misclassifying class 2 as class 1 is greater than class 1 as class 2.

$$\frac{p(x\,|\omega_1)}{p(x\,|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} * \frac{P(\omega_2)}{P(\omega_1)}$$

λ_{12} is the penalty associated with misclassifying class 1 as class 2

λ_{21} is the penalty associated with misclassifying class 2 as class 1

λ_{11} *and* λ_{22} are assumed to be 0 because there is no penalty for correct classification.

In this case, we assume that $\lambda_{21} > \lambda_{12}$. We know that both probabilities $P(\omega_2)$ and $P(\omega_1)$ are 0.5, the $\frac{P(\omega_2)}{P(\omega_1)}$ term can be neglected. The optimal threshold is when the ccp for both classes are equal ($\frac{p(x\,|\omega_1)}{p(x\,|\omega_2)}$ =1). The threshold value can be defined as:

$$\frac{p(x\,|\omega_1)}{p(x\,|\omega_2)} = \frac{\lambda_{12}}{\lambda_{21}}$$

And since $\lambda_{21} > \lambda_{12}$, we can see that

$$\frac{p(x\,|\omega_1\,)}{p(x\,|\omega_2\,)} < 1$$

Finally, we can conclude that:

$$p(x\,|\omega_1) < p(x\,|\omega_2)$$

So in this case, the threshold line will move towards the left where $p(x\,|\omega_2)$ is bigger.

6. Adjust your program to accept Sepal Length as the discriminating feature g(x1). Suggest which of the two features (x1, x2) might be a better choice for separating the two classes w_1 and w_2. Justify.

Bayes decision theory involves computation of posterior probabilities and the decision regarding which class an unknown object belongs to is based upon which posterior probability is the highest. However, we also need to consider the error rate. This error rate equates to the sum of all posterior probabilities that we did not choose. For example, for a two categories dichotomizer, if we choose *ω1*, then the error rate will be equal to *P(ω2 | x)*.

Table 2: Classifying samples by label, posterior, and discriminate function according to sepal length

X_1	Class	Posterior Probabilities $[P(w_1\|x),P(w_2\|x)]$	Discriminate Function [g(x)]
3.3	w1	[0.7204, 0.2796]	0.4408
4.4	w1	[0.9288, 0.0712]	0.8576
5.0	w1	[0.7795, 0.2205]	0.5589
5.7	w2	[0.0984, 0.9016]	-0.8032
6.3	w2	[0.0010, 0.9990]	-0.9979

Sepal length is the better choice for separating the two classes as the variance of the sepal length is greater than the variance of the sepal width. This means that the lengths are more spread out than the width which means there is less overlap between the samples if you classify them by length. Since, we are minimizing the classification error, the accuracy is improved.