# UNIVERSITI KUALA LUMPUR
# MALAYSIAN INSTITUTE OF INDUSTRIAL TECHNOLOGY

## FINAL EXAMINATION

## JANUARY 2016 SEMESTER

| | | |
|---|---|---|
| COURSE CODE | : | JCB 30903 |
| COURSE TITLE | : | PATTERN RECOGNITION AND CLASSIFICATION |
| PROGRAMME LEVEL | : | BACHELOR |
| DATE | : | 18 MAY 2016 |
| TIME | : | 2.30 PM – 5.30 PM |
| DURATION | : | 3 HOURS |

### INSTRUCTIONS TO CANDIDATES

1. Please read the instructions given in the question paper CAREFULLY.
2. This question paper is printed on both sides of the paper.
3. This question paper consists of TWO (2) sections.
4. Answer ALL questions in Section A. Choose ONE (1) questions in section B.
5. Please write your answers on the answer booklet provided.
6. Please answer all questions in English only.

THERE ARE 5 PAGES OF QUESTIONS EXCLUDING THIS PAGE.

**SECTION A (Total: 75 marks)**

**INSTRUCTION: Answer ALL questions.**
**Please use the answer booklet provided.**

**Question 1**

To assess the efficacy of speed limit systems on Malaysian highways, Malaysian authorities have measured the driver's mortality rates associated with vehicles' weight and speed involved in head-on collisions as shown in Figure 1. Data collected from 10 different accident cases is shown in Table 1.

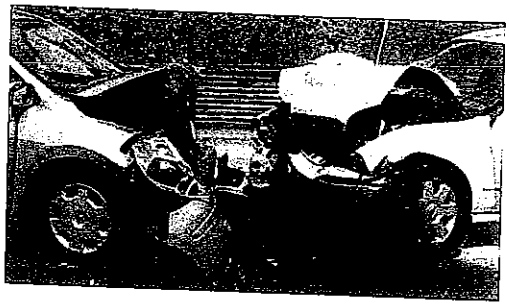


Figure 1: Head-on Collision

Table 1: Mortality rate for 10 accident cases

| | Speed (km/h) | Weight (Kg) | Survival Rate |
|---|---|---|---|
| Case 1 | 135 | 1335 | Dead |
| Case 2 | 128 | 1456 | Dead |
| Case 3 | 78 | 2879 | Alive |
| Case 4 | 88 | 2340 | Alive |
| Case 5 | 116 | 1881 | Dead |
| Case 6 | 122 | 1862 | Alive |
| Case 7 | 127 | 1463 | Dead |
| Case 8 | 100 | 2328 | Alive |
| Case 9 | 110 | 2183 | Alive |
| Case10 | 112 | 2073 | Dead |

A pair of speed cameras and weight recorders for vehicles has been installed in Tangkak. A car weighing 2000 kg has been through the area with a speed of 116 km/h. Predict the survival rate of the driver in the event of collision with a vehicle ahead using Bayes classifier and determine the accuracy of the decision

Given:

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{and} \quad p(x_1,x_2,\ldots x_i \mid c_j)=\frac{1}{\sqrt{\sum\left|(2\pi)^d\right.}}e^{\left(-\frac{1}{2}(x-\mu)^T\sum^{-1}(x-\mu)\right)}$$

(25 marks)

**Question 2**

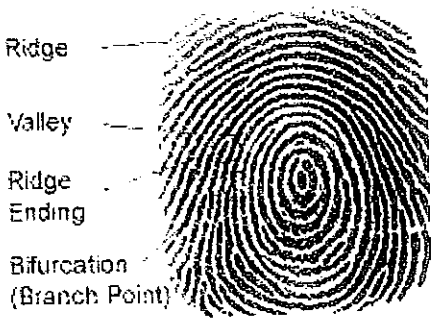



Figure 2 : Fingerprint

Comparison between finger sizes and the number of its ridges bifurcation for 10 different students is given as in Table 2:

Table 2: Comparison between finger sizes and number of ridges

| | Area ($cm^2$) | Number of ridges |
|---|---|---|
| Student 1 | 1.74 | 131 |
| Student 2 | 1.79 | 133 |
| Student 3 | 2.34 | 137 |
| Student 4 | 2.40 | 137 |
| Student 5 | 2.56 | 145 |
| Student 6 | 2.62 | 146 |
| Student 7 | 2.71 | 146 |
| Student 8 | 2.94 | 151 |
| Student 9 | 3.06 | 154 |
| Student 10 | 3.46 | 157 |

a) Choose the Eigen vector that shows the most discriminative features from the above data.

(18 marks)

b) Construct the new data by considering only the highest principle component values.

(2 marks)

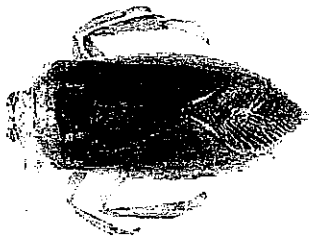
c) Compare the new data and the original data and determine the absolute mean error between the data.

(5 marks)

**Question 3**

An engineer has been asked to develop a system to differentiate between Cockroaches and Water Bugs. The engineer uses the length and grayscale color intensity of these insects as features as shown in Figure 3.

Cockroach



Water Bug

Figure 3 : Cockroaches and Water Bugs features

Data collected from 10 insects is shown in Table 3:

Table 3 : Insects features

| | Length (mm) | Gray color intensity | Insect |
|---|---|---|---|
| Insect 1 | 43 | 37 | Water Bug |
| Insect 2 | 57 | 50 | Cockroach |
| Insect 3 | 38 | 50 | Water Bug |
| Insect 4 | 49 | 68 | Water Bug |
| Insect 5 | 34 | 74 | Water Bug |
| Insect 6 | 48 | 78 | Cockroach |
| Insect 7 | 45 | 84 | Cockroach |
| Insect 8 | 42 | 96 | Water Bug |
| Insect 9 | 53 | 132 | Cockroach |
| Insect 10 | 52 | 142 | Cockroach |

If a new image from an unknown insect with length = 20 mm and gray color intensity=140, classify the type of insect for the unknown image using **LINEAR DISCRIMINANT ANALYSIS** by reducing the dimension of the data.

(25 marks)

**SECTION B (Total: 15 marks)**

**INSTRUCTION : Answer ONE (1) question only.**
**Please use the answer booklet provided.**

**Question 1**

Marks of the following three courses tabulated in Table 4 have been used by ICE academic advisors in advising their advisee on the best elective subject to be taken in semester 4. Predict the elective subject required to be taken by Ahmad using **K-NEAREST NEIGHBOR (KNN)** based on the given dataset.

Table 4 : Historical marks prediction for an elective subject

| State | JCB 20103 | JCB 20203 | JCB 20303 | Subject |
|---|---|---|---|---|
| Student 1 | 43 | 65 | 71 | JCB 41843 |
| Student 2 | 50 | 44 | 70 | JCB 41843 |
| Student 3 | 70 | 18 | 86 | JCB 30903 |
| Student 4 | 60 | 37 | 81 | JCB 30903 |
| Student 5 | 80 | 38 | 84 | JCB 30903 |
| Student 6 | 55 | 44 | 77 | JCB 30903 |
| Student 7 | 67 | 44 | 66 | JCB 41843 |
| Student 8 | 55 | 48 | 73 | JCB 41843 |
| Student 9 | 70 | 51 | 82 | JCB 30903 |
| Student 10 | 51 | 55 | 64 | JCB 41843 |
| Student 11 | 51 | 58 | 75 | JCB 41843 |
| Student 12 | 74 | 61 | 82 | JCB 30903 |
| Student 13 | 72 | 63 | 86 | JCB 30903 |
| Student 14 | 53 | 65 | 75 | JCB 41843 |
| **Ahmad** | **60** | **50** | **75** | **?** |

Construct KNN classifier using three different distance measures namely Sorensen Distance, Cosine Distance and City Block Distance to predict which elective subject Ahmad should enrolled. Use k=7.

(15 marks)

**Question 2**

A sport reporter has conducted a study regarding JDT soccer match in order to predict the performance of future games either Win, Lose or Draw, gathered from past 10 games. The data has 4 attributes which are the location of the games (Location), starting time for the match (Time), moral spirits of the player before starting the game (Moral State) and the number of key players unable to play due to injuries (Injuries). The result of the study is given in Table 5.

Table 5 : JDT soccer match performance

| Location | Time | Moral Spirit | Injuries | Outcome |
|---|---|---|---|---|
| Home | 5:00 PM | Low | 0 | Won |
| Home | 8:00 PM | Low | 1 | Won |
| Away | 5:00 PM | Low | 0 | Won |
| Home | 8:00 PM | Low | 1 | Won |
| Away | 8:00 PM | High | 2 | Draw |
| Home | 10:00 AM | High | 1 | Draw |
| Away | 10:00 AM | High | 2 | Draw |
| Away | 8:00 PM | Low | 1 | Lost |
| Home | 5:00 PM | Medium | 1 | Lost |
| Away | 8:00 PM | Medium | 1 | Lost |

Suppose there are 3 more games for JDT schedule as follows:

| Date | Location | Time | Moral Spirit | Injuries |
|---|---|---|---|---|
| 5 Feb 2015 | Home | 8:00 PM | Medium | 2 |
| 14 Feb 2015 | Home | 5:00 PM | Low | 1 |
| 28 Feb 2015 | Away | 8:00 PM | Low | 0 |

Construct **DECISION TREE** classifier to predict the outcome of the soccer game either win, lose or draw. Use Entropy to calculate the information gain for the decision tree.

(15 marks)

**END OF EXAMINATION PAPER**