

UNIVERSITI MALAYA
UNIVERSITY OF MALAYA

PEPERIKSAAN IJAZAH SARJANA MUDA SAINS KOMPUTER
EXAMINATION FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE

SESI AKADEMIK 2021/2022 : SEMESTER II
ACADEMIC SESSION 2021/2022 : SEMESTER II

WIA1006/WID3006 : Pembelajaran Mesin
Machine Learning

Jun/Julai 2022
June/July 2022

Masa: 72 jam
Time: 72 hours

ARAHAN KEPADA CALON:
INSTRUCTIONS TO CANDIDATES:

Calon dikehendaki menjawab **SEMUA** soalan (100 markah).
Answer **ALL** questions (100 marks).

(Kertas soalan ini mengandungi 4 soalan dalam 7 halaman yang dicetak)
(This question paper consists of 4 questions on 7 printed pages)

1. Answer all questions (25 marks).

a) Define the following terms and state two industrial applications for each learning algorithm.

- i. Supervised learning
- ii. Unsupervised learning

(3 marks)

b) One of the examples of unsupervised learning is clustering, where K-means is an iterative clustering algorithm that aims to assign data points into K groups, and K represents the number of clusters. Answer following questions related to K-means algorithm,

- i. State a method that is commonly used to choose the number of clusters, K.

(1 mark)

- ii. Explain the steps of the algorithm and state the stopping criteria.

(6 marks)

c) State two benefits of performing dimensionality reduction.

(2 marks)

d) One of the examples of dimensionality reduction is PCA algorithm. Specify the steps of this algorithm.

(3 marks)

Table 1

x	y
0.5	0.7
1.0	1.1
1.5	1.6
1.1	0.9
1.9	2.2
2.2	2.9
2.6	2.7
2.0	1.6
2.5	2.4
3.1	3.0

- e) By referring a set of data provided in Table 1, develop the Python code to perform dimensionality reduction using PCA. algorithm and compute the followings:

i. Feature vector with both eigenvectors.

(2 marks)

ii. New data set (from 2D to 1D reduction)

(8 marks)

2. For the following questions, provide the justification for your answer in a brief and concise manner. If answering using examples is necessary, please provide them in your answer.

Answer all questions (25 marks).

- a) You design a fully connected neural network architecture where all activations are sigmoid. Should you initialize the weights with large or small positive numbers? How will these affect your model?

(3 marks)

- b) You are given a dataset of 10×10 grayscale images. Your goal is to build a 5-class classifier. You have to adopt one of the following two options:

- the input is flattened into a 100-dimensional vector, followed by a fully-connected layer with 5 neurons
- the input is directly given to a convolutional layer with five 10×10 filters

Explain which one you would choose and why.

(3 marks)

- c) You are doing full batch gradient descent using the entire training set (not stochastic gradient descent). Is it necessary to shuffle the training data? Justify your answer.

(3 marks)

- d) One of your friends has trained a cat vs. non-cat classifier. It performs very well and you want to use transfer learning to build your own model. Explain what additional hyperparameters (due to the transfer learning) you will need to tune.

(3 marks)

e) Given below is the designed pipeline for autonomous driving system:

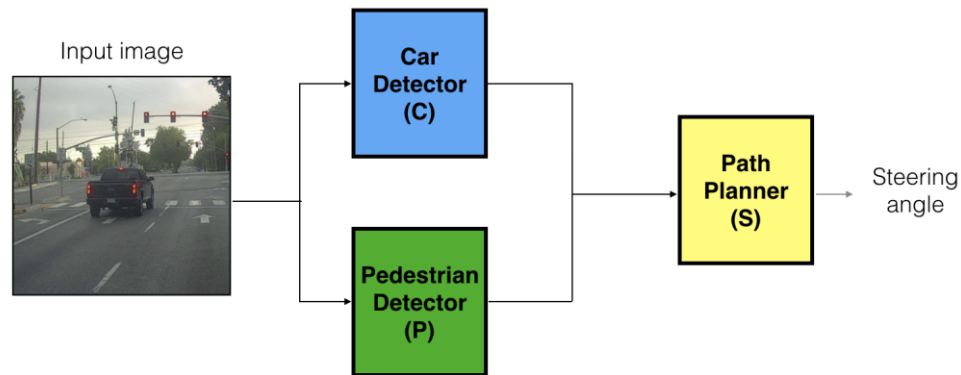


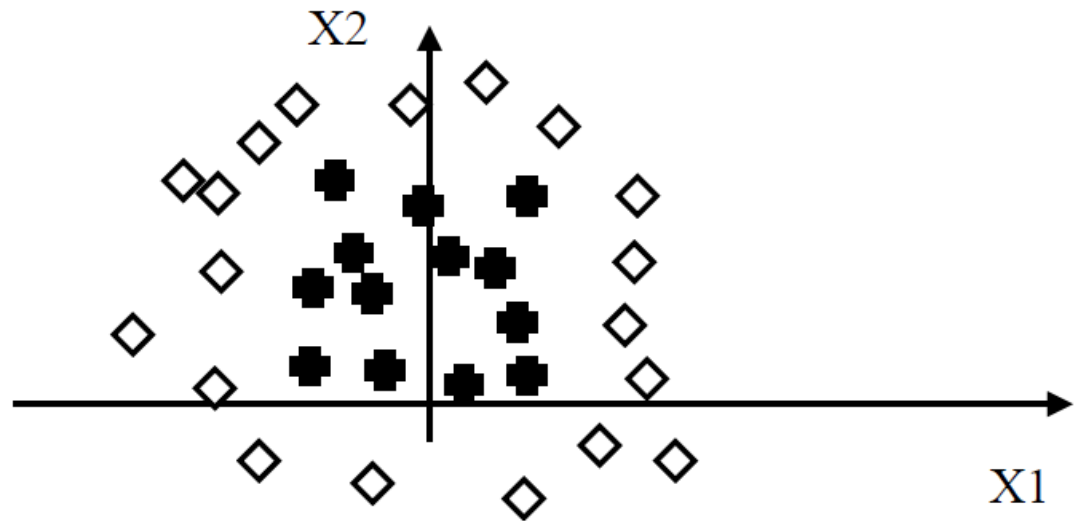
Figure 1: The input camera image is given to two modules: the Car Detector C and the Pedestrian Detector P. C outputs a set of bounding boxes localizing the cars. P outputs a set of bounding boxes localizing the pedestrians. The bounding boxes are then given to a Path Planner S which outputs the steering angle. Assume all these submodules are supervised learning algorithms.

- i. What data do you need to train the submodules in the pipeline presented in Figure 1?
(3 marks)
 - ii. Explain how would you collect the data for the 3 modules; (X_c, Y_c) , (X_p, Y_p) and (X_s, Y_s) , referring to Car Detector, Pedestrian Detector and Path Planner respectively.
(6 marks)
 - iii. After setting up a camera on the hood of your car, an expert driver in your team has collected the dataset A in a city where most roads are straight. She always drives in the center of the lane. Describe two problems that can occur if you train a model on A and test it on real-world data.
(4 marks)
3. Answer all questions (25 marks).
- a) The Malaysian Railway Consortium have been trialling 2 different machine learning methods which attempt to predict whether a train will arrive at its destination on time or not, using several input features corresponding to weather conditions, train priorities, ongoing repair works etc. (for this purpose, 'on time' is defined as the train arriving no more than 10 minutes after its scheduled time).

Two different machine learning methods have been tested on a common set of 500 train runs, and the results are as follows:

Method	Actually on time	Actually late	Total predictions
Method 1 predicted on time	131	155	286
Method 1 predicted late	19	195	214
Method 2 predicted on time	82	72	154
Method 2 predicted late	68	278	346

- i. Compute the F_1 score for Method 1 and Method 2 for the trains being on time.
(4 marks)
 - ii. Based on the F_1 score for Method 1 and Method 2, determine which of the method is the best method for the task.
(1 mark)
 - iii. Explain the concept of precision and recall using the train prediction problem.
(7 marks)
- b) Suppose that we want to build a SVM classifier that classifies two-dimensional data (i.e., $X = [x_1, x_2]$) into two classes: diamonds and crosses. We have a set of training data that is plotted as follows:



Explain how you would build the SVM classifier with respect to the classification problem provided by providing detailed explanations for the followings:

- i. The support vector optimization function used to classify the two different classes.

Is there a need to use a kernel function to solve the problem? Why? And which kernel is most suitable? Explain.

(8 marks)

- ii. The concept of large margin classifier with regards to SVM and the problem presented.

(2 marks)

- iii. What is the C parameter in SVM, and how will it [affect](#) the final classification result for the given problem?

(3 marks)

4. Answer all questions (25 marks).

- a) Can the following boolean function represented with a single logistic threshold unit (i.e., a single unit from a neural network)? If yes, show the weights. If not, explain why not.

A	B	$f(A,B)$
1	1	0
0	0	0
1	0	1
0	1	0

(5 marks)

- b) Draw the single logistic threshold unit for the input and output as shown in the table. Indicate the input, the weights and the output.

(5 marks)

- c) Assume we have a set of data from patients who have visited Universiti Malaya Medical Centre (UMMC) during the year 2022. A set of features (e.g., temperature, height) have been also extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases). We have decided to use a neural network to solve this problem. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer.

(5 marks)

- d) Does increasing the number of hidden nodes in a multilayer perceptron improve generalisation? Why (not)?

(5 marks)

- e) When training perceptrons with gradient descent, one can add momentum: why and how does momentum help?

(5 marks)