THE UNIVERSITY OF SYDNEY

CONFIDENTIAL EXAM PAPER

This paper must not be circulated in any way and must not be removed from the exam venue

School of Computer Science

EXAMINATION

Semester 1 - Main, 2021

COMP5318 Machine Learning and Data Mining

EXAM WRITING TIME: 2 hours

READING TIME: 10 minutes

EXAM CONDITIONS:
This is a RESTRICTED OPEN book examination - specified materials permitted.

All submitted work must be completed individually without consulting anybody else, without browsing the internet or using other materials and devices apart from the permitted, in accordance with the University Policy on "Academic Honesty in Coursework". All submissions will go through TurnItIn for plagiarism detection and the penalties are severe.

MATERIALS PERMITTED IN THE EXAM:

1. Teaching materials from this course – lecture slides and tutorial notes
2. One page of student's own notes - double-sided A4 size, handwritten or typed. This page must be uploaded before the exam to the Canvas COMP5318 website.
3. Calculator – non-programmable

MATERIALS TO BE SUPPLIED TO STUDENTS: None

INSTRUCTIONS TO STUDENTS:

1. Type your answers in your text editor (Word, Latex, etc), convert the file into a pdf file and submit it to Canvas. No other file format will be accepted.
2. Hand-written responses will not be accepted, you need to type your answers.
3. Start by typing your student ID number on the first page. Do not type your name as the marking is anonymous.
4. Submit only your answers to the questions, do not copy the questions.

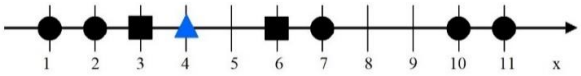
For examiner use only:

Q1 (13)	Q2 (10)	Q3 (10)	Q4 (13)	Q5 (14)	Q6 (11)	Q7 (12)	Q8 (17)	Total (100)

Question 1 [13 marks]

Select the correct answer and provide a brief explanation:

1. [2 marks] The figure below shows a training set of 8 examples described with one numerical feature x and belonging to two classes: circles and squares. A new example is shown with a blue triangle.

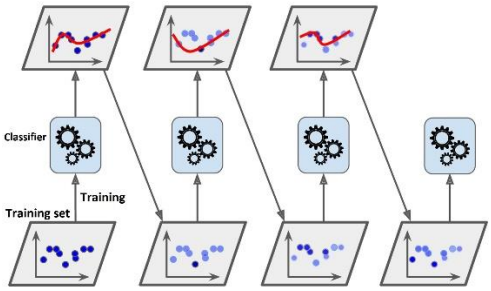


What will be the prediction of **3-Nearest Neighbor** for the class of the new example? If there are ties, settle them by choosing the example on the left.

Circle Square

Explanation:

2. [2 marks] Does this diagram correspond to Bagging or Boosting?



Bagging Boosting

Explanation:

3. [3 marks] The kernel trick in support vector machines ensures that the data will be linearly separable in the new space.

True False

Explanation:

4. [3 marks] Given is the following training data, where *occupation*, *age* and *loan-salary-ratio* are the features and *outcome* is the class. Two prediction models are built, Model 1 and Model 2, both consistent with the training data.

occupation	age	loan-salary-ratio	outcome
industrial	39	3.40	default
industrial	22	4.02	default
professional	30	2.70	repay
professional	27	3.32	default
professional	40	2.04	repay
professional	50	6.95	default
industrial	27	3.00	repay
industrial	33	2.60	repay
industrial	30	4.50	default
professional	45	2.78	repay

Model 1:

if loan-salary-ratio > 3.00 then outcome = default
else outcome = repay

Model 2:

if age = 50 then outcome = default
else if age = 39 then outcome = default
else if age = 30 and occupation = industrial then outcome = default
else if age = 27 and occupation = professional then outcome = default
else outcome = repay

Which of these two models is more likely to generalize better on new examples?

Model 1 Model 2

Explanation:

5. [3 marks] At each step, PRISM selects the best attribute by considering all classes.

True False

Explanation:

Question 2 [10 marks]

Given is the following training data, where *city* and *season* are the features and *price* is the class:

city	**season**	**price**
Madrid	summer	high
Barcelona	spring	medium
Madrid	spring	medium
Barcelona	summer	high
Bilbao	winter	medium
Sevilla	spring	high
Sevilla	winter	medium
Bilbao	summer	medium

Use Naïve Bayes to predict the value of *price* for the following new example: *city=Sevilla, season=summer*. Show your calculations.

THE UNIVERSITY OF SYDNEY

Question 3 [10 marks]

Given is the following training data, where *restaurant* and *time* are the features and *price* is the class:

restaurant	**time**	**price**
casual	dinner	high
casual	lunch	medium
family	lunch	medium
family	dinner	high
cafe	lunch	high
cafe	breakfast	medium
fast	breakfast	medium
fast	dinner	medium

You may use this table:

x	y	$-(x/y)*\log_2(x/y)$	x	y	$-(x/y)*\log_2(x/y$
1	2	0.50	1	7	0.40
1	3	0.53	2	7	0.52
2	3	0.39	3	7	0.52
1	4	0.5	4	7	0.46
3	4	0.31	5	7	0.35
1	5	0.46	6	7	0.19
2	5	0.53	1	8	0.38
3	5	0.44	3	8	0.53
4	5	0.26	5	8	0.42
1	6	0.43	7	8	0.17
5	6	0.22			

a) What is the entropy of this data set with respect to the class?

b) What is the information gain of *restaurant*? Show your calculations.

Question 4 [13 marks]

1. [4 marks] There are 100 students in a computer science course. Isabella consistently outperforms the other students on the assessments during the semester and on the final exam he gets a mark of 99 while the next highest mark is 75. The range of exam marks is between 5 and 99. We would like to fit a **linear regression** model to the exam marks. Would Isabella's mark cause problems? Briefly explain your answer.

2. [2 marks] List one advantage of **Lasso regression** compared to the **standard linear regression** and briefly explain your answer.

3. [2 marks] In **random forest**, how is the correlation among the combined decision trees reduced?

4. [5 marks] Consider the task of predicting credit card fraud in real time using a machine learning classifier. This task requires that the classifier performs thousands of predictions per second. Which algorithm is more suitable: **k-nearest neighbor** or **logistic regression**? Explain your answer.

Question 5 [14 marks]

1. [8 marks] A company is building a classifier to predict if customers will like new products. The classifier takes as an input a vector with a very high dimensionality, has to be trained on a very large dataset and also has to be updated frequently and efficiently as new data comes in. Which of the following machine learning algorithms is most appropriate? Explain your answer.

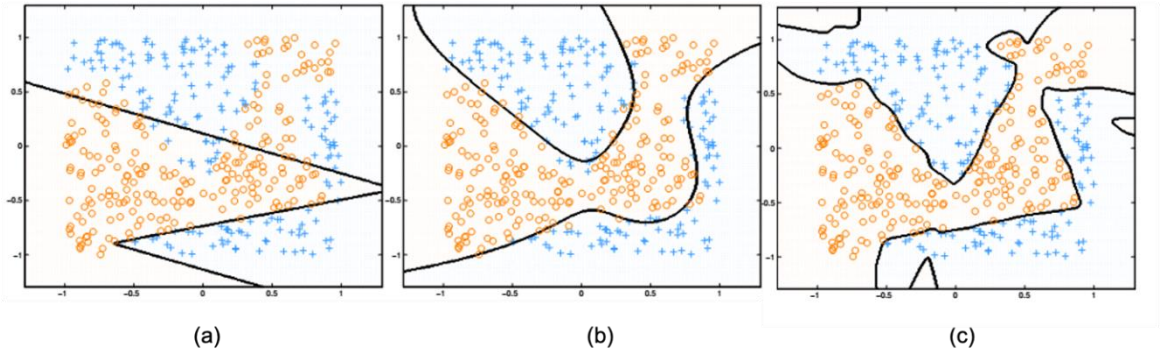
1. Nearest neighbor
2. Decision tree
3. Naïve Bayes
4. Support vector machine

2. [6 marks] Bob used Principal Component Analysis (PCA) for dimensionality reduction in a classification task. The data consisted of *n* examples, *m* features and 2 classes. Each example corresponded to a patient's record, each attribute represented a patient's symptom and the two classes corresponded to presence or absence of a disease. The *n x m* dimensional data was transformed using SVD. Bob explained that instead of using all *m* features, we can use the top *k* principal components as an input to the classifier. To find the *k* components he computed the correlation of each feature with the class value and selected the *k* features with highest correlation. Identify Bob's misconceptions about PCA and correct them.

Question 6 [11 marks]

1. [4 marks] Is it a good idea to initialize the weights of a deep neural network to zero? Explain your answer.

2. [3 marks] Consider the decision boundaries in the figure below. Which one could be the result of using a deep neural network?



A. (a)
B. (b)
C. (c)
D. (a) and (b)
E. (a), (b) and (c)

Explanation:

3. [4 marks] For what type of problems is DBSCAN more suitable than k-means?

Question 7 [12 marks]

1. [4 marks] In a convolutional neural network layer, we use the input x and the convolutional filter w as shown in the figure below. The convolutional stride is 2 and the padding size (using zero padding) is 1.

3	1	-2	0	4
0	2	0	-2	1
2	1	-1	0	0
1	-2	3	1	1
2	0	2	-1	3

x

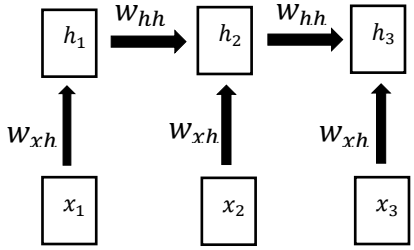
0	-1	0
-1	0	1
0	1	1

w

a) What will be the dimensionality of the feature map resulting from the convolution operation?

b) Calculate the first two values of this feature map and briefly show your calculations.

2. [6 marks] Given is the following recurrent neural network. The activation functions are identity functions (the input is the same as the output) and the initial $h_0 = \vec{0}$.



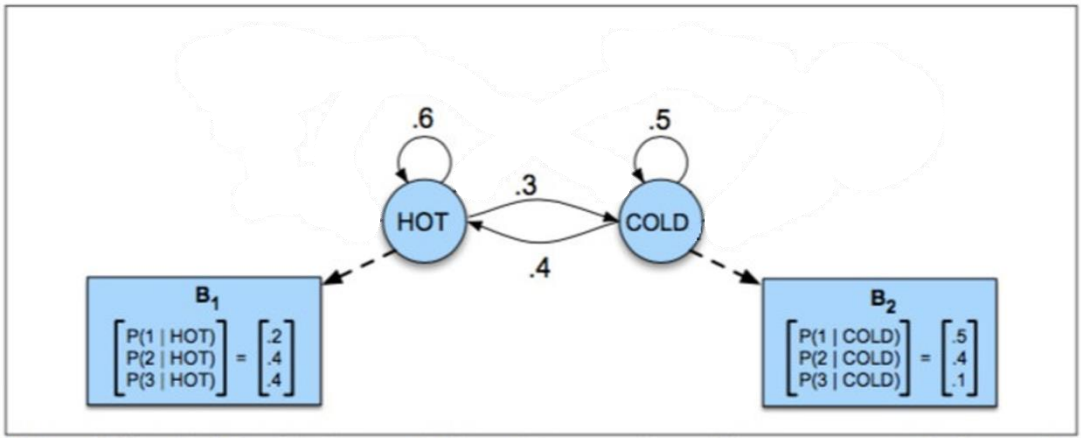
Given: $x_1 = [1, 2], x_2 = [1, 1], w_{xh} = \begin{bmatrix} 0.5 & 0.7 & -0.5 \\ 0.5 & 0.8 & 0.25 \end{bmatrix}^T$ and $w_{hh} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix}^T$, what will be the hidden states of h_1 and h_2? Briefly show your calculations.

3. [2 marks] List one difference between reinforcement learning and supervised learning and explain it.

Question 8 [17 marks]

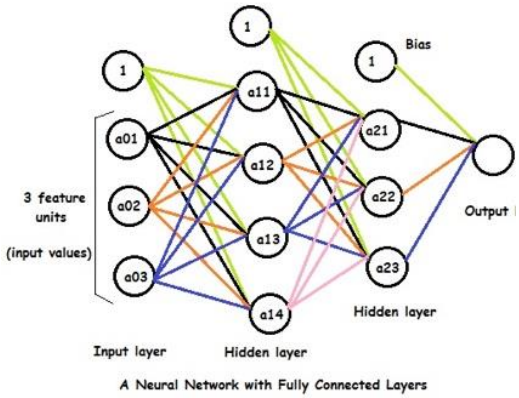
1. [9 marks] The diagram below shows the Hidden Markov Model for this scenario:

Given is a sequence of observations, where each observation is an integer representing the number of ice creams Jason has eaten on a given day. Find the hidden sequence of weather states (*hot* or *cold*) which caused Jason to eat the ice cream.



Suppose you know that Jason ate 1, 2 and 3 ice creams respectively during the last three days. Unfortunately, you do not know the weather states of the last two days but you know that the weather state of the first day (when Jason ate 1 ice cream) was *hot*. Specify all possible weather sequences for the last three days and explain which sequence is most likely. Show your calculations.

2. [8 marks] Given is a dataset $\mathcal{D}$, where the i-th example is a 4-dimensional vector $\boldsymbol{x^i} = \left[x_1^i, x_2^i, x_3^i, x_4^i\right]^T$ and y^i is the associated target value. Professor X asked Bob to use the neural network architecture given below for this prediction task.



a) Since there are 3 inputs in this architecture, Bob needs to reduce the dimension of the data to 3. How can this data reduction task be accomplished by using a neural network?

b) Bob also wants to try Principal Component Analysis (PCA) to do the dimensionality reduction. What is the difference between your method from a) above and PCA?

c) Professor X told Bob that there is redundant information in the third and fourth dimension of each example. Hence, for each example $\boldsymbol{x}^{\boldsymbol{i}}$, Bob should keep the first two dimensions, while for the last two dimensions, it is better to learn an **agent** to determine which one to keep as the last input entry for the network. In a potential **reinforcement learning** solution, how Bob should define the state, action and reward for the agent?

END OF EXAMINATION