60

(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS
2019/2020 ACADEMIC YEAR

END OF SEMESTER EXAMINATIONS
YEAR FOUR SEMESTER TWO

FOR THE DEGREE OF
COMPUTER SCIENCE

COURSE CODE : CSC 423

COURSE TITLE : MACHINE LEARNING

DATE: 12/11/2020 TIME: 9.00 A.M – 11.00 A.M

INSTRUCTIONS:

ANSWER QUESTIONS ONE AND ANY OTHER TWO

QUESTION ONE [COMPULSORY] [30 MARKS]

a) A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. What would be a reasonable choice for T and for P? [4 marks]

b) The amount of rain that falls in a day is usually measured in either millimeters (mm) or inches. Suppose you use a learning algorithm to predict how much rain will fall tomorrow. Would you treat this as a classification or a regression problem and why? [2 marks]

c) Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will win a patent infringement lawsuit (by training on data of companies that had to defend against similar lawsuits). Would you treat this as a classification or a regression problem and why? [2 marks]

d) Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. For each of the following state with reasons weather you would apply supervised learning or unsupervised learning? In each case, assume some appropriate dataset is available for your algorithm to learn from. [8 marks]

I. Given data on how 1000 medical patients respond to an experimental drug (such as effectiveness of the treatment, side effects, etc.), discover whether there are different categories or "types" of patients in terms of how they respond to the drug, and if so what these categories are.

II. Given a large dataset of medical records from patients suffering from heart disease, try to learn whether there might be different clusters of such patients for which we might tailor separate treatments

III. Have a computer examine an audio clip of a piece of music, and classify whether or not there are vocals (i.e., a human voice singing) in that audio clip, or if it is a clip of only musical instruments (and no vocals).

IV. Given genetic (DNA) data from a person, predict the odds of him/her developing diabetes over the next 10 years.

e) Give an alternative definition of machine learning other than the one given in question 1.

[2 marks]

f) Consider the problem of predicting how well a student does in her second year of university, given how well she did in her first year. Specifically, let **x** be equal to the number of "A" grades that a student receives in their first year of college. We would like to predict the value of **y**; which we define as the number of "A" grades they get in their second year.

In linear regression, our hypothesis $h_\theta(x)=\theta_0+\theta_1 x$ and we use **m** to denote the number of training examples.

x	y
5	4
3	4
0	1
4	3

For the training set given above, what is the value of **m** the number of training examples?

[1 mark]

g) Suppose we set $\theta_0=-2,\theta_1=0.5$ in the linear regression hypothesis from Q1(f). What is $h\theta$ (6)?

[2 marks]

h) Consider the following training set of 4 training examples:

x	y
1	0.5
2	1
4	2
0	0

Consider the linear regression model $h_\theta(x) =\theta_0+\theta_1 x$. What are the values of θ_0 and θ_1 that you would expect to obtain upon running gradient descent on this model? (Linear regression will be able to fit this data perfectly.) [2 marks]

i) Briefly describe one supervised algorithm and one unsupervised algorithm of your choice.

[7 marks]

QUESTION TWO [20 MARKS]

a) What is the fundamental idea behind Support Vector Machines? [3 marks]

b) What is a support vector? [2 marks]

c) Why is it important to scale the inputs when using SVMs? [5 marks]

d) How should you set the QP parameters (**H**, **f**, **A**, and **b**) to solve the soft margin linear SVM classifier problem using an off-the-shelf QP solver? [10 marks]

QUESTION THREE [20 MARKS]

a) If a Decision Tree is overfitting the training set, state with reasons whether it is a good idea to try decreasing max_depth? [3 marks]

b) If a Decision Tree is underfitting the training set, state with reasons whether it is a good idea to try scaling the input features? [3 marks]

c) If it takes one hour to train a Decision Tree on a training set containing 1 million instances, roughly how much time will it take to train another Decision Tree on a training set containing 10 million instances? [4 marks]

d) What is the approximate depth of a Decision Tree trained (without restrictions) on a training set with 1 million instances? [4 marks]

e) Is a node's Gini impurity generally lower or greater than its parent's? Is it *generally* lower/greater, or *always* lower/greater? [6 marks]

QUESTION FOUR [20 MARKS]

a) Define ensemble learning. [2 marks]

b) If you have trained five different models on the exact same training data, and they all achieve 95% precision, is there any chance that you can combine these models to get better results? If so, how? If not, why? [5 marks]

c) Distinguish between hard and soft voting classifiers? [2 marks]

d) If your Gradient Boosting ensemble overfits the training set, should you increase or decrease the learning rate? [3 marks]

e) What is the benefit of out-of-bag evaluation? [2 marks]

f) What makes Extra-Trees more random than regular Random Forests? How can this extra randomness help? Are Extra-Trees slower or faster than regular Random Forests? [6 marks]

QUESTION FIVE [20 MARKS]

a) What are the main motivations for reducing a dataset's dimensionality? What are the main drawbacks? [3 marks]

b) What is the curse of dimensionality? [2 marks]

c) Once a dataset's dimensionality has been reduced, is it possible to reverse the operation? If so, how? If not, why? [6 marks]

d) State with reasons whether PCA can be used to reduce the dimensionality of a highly nonlinear dataset. [3 marks]

e) Suppose you perform PCA on a 1,000-dimensional dataset, setting the explained variance ratio to 95%. How many dimensions will the resulting dataset have? Show how you arrive at your answer. [6 marks]