[Knowledge for Development]

KIBABII UNIVERSITY

[KIBU]

UNIVERSITY EXAMINATIONS
2019/2020 ACADEMIC YEAR

SPECIAL/SUPPLEMENTARY EXAMINATIONS
YEAR FOUR SEMESTER ONE EXAMINATIONS

FOR THE DEGREE IN
COMPUTER SCIENCE

COURSE CODE : CSC 423.

COURSE TITLE : MACHINE LEARNING

DATE: 15/02/2021   TIME: 2:00 P.M – 4:00 A.M.

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO

QUESTION ONE [COMPULSORY] [30 MARKS]

a) Define Machine Learning and name four types of problems where it shines. [5 marks]
b) What are the two most common supervised tasks? [2 marks]
c) Name four common unsupervised tasks? [4 marks]
d) 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]
e) What type of Machine Learning algorithm would you use to:. [4 marks]
   i. allow a robot to walk in various unknown terrains?
   ii. segment your customers into multiple groups?
f) Would you frame the problem of email spam detection as a supervised learning problem or an unsupervised learning problem? Give reasons. [4 marks]
g) Outline any four of the main challenges in Machine Learning. [4 marks]
h) If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name three possible solutions? [5 marks]

QUESTION TWO [20 MARKS]

a) Suppose you use Batch Gradient Descent and you plot the validation error at every epoch. If you notice that the validation error consistently goes up, what is likely going on? How can you fix this? [4 marks]
b) Suppose you are using Polynomial Regression. You plot the learning curves and you notice that there is a large gap between the training error and the validation error. What is happening? What are three ways to solve this? [4 marks]
c) Suppose the features in your training set have very different scales. What algorithms might suffer from this, and how? What can you do about it? [4 marks]
d) You are given the task of classifying pictures as outdoor/indoor and daytime/nighttime. State with reasons whether you should implement two Logistic Regression classifiers or one Softmax Regression classifier. [4 marks]

QUESTION THREE [20 MARKS]

a) What Linear Regression training algorithm can you use and why, if you have a training set with millions of features? [4 marks]
b) Can Gradient Descent get stuck in a local minimum when training a Logistic Regression model? Why? [4 marks]
c) Why is it important to scale the inputs when using SVMs? [3 marks]
d) State with reasons whether you should use the primal or the dual form of the SVM problem to train a model on a training set with millions of instances and hundreds of features. [5 marks]

e) What is a test set and why would you want to use it? [4 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 FIVE [20 MARKS]

a) Suppose we train a hard-margin linear SVM on $n > 100$ data points in R2, yielding a hyperplane with exactly 2 support vectors. If we add one more data point and retrain the classifier, what is the maximum possible number of support vectors for the new hyperplane (assuming the $n+1$ point are linearly separable)? [3 marks]

b) Suppose we are given data comprising points of several different classes. Each class has a different probability distribution from which the sample points are drawn. We do not have the class labels. We use k-means clustering to try to guess the classes. What circumstances would undermine its effectiveness? [4 marks]

c) You've just finished training a random forest for spam classification, and it is getting abnormally bad performance on your validation set, but good performance on your training set. Your implementation has no bugs. Give four possible causes of the problem? [4 marks]

d) In terms of the bias-variance decomposition, a 1-nearest neighbor classifier has _____ than a 3-nearest neighbor classifier. [4marks]

e) What is the primary motivation for using the kernel trick in machine learning algorithms? [2 marks]

f) Give two examples of a loss function for classification problems. [3 marks]