CS3244 Exam 1: Part 1

28 Sep 2020

Please do not turn to the next page until you are told to do so by your proctor.

- This exam part is worth **25** marks out of a **75** mark total for all three parts.
- This exam part is estimated to take you about **25** minutes to complete.
- This exam part has a total of **13** questions.
- This exam part contains only multiple choice questions (MCQs) and multiple response questions (MRQ). Please key these into the assessment system.
- You can visit `http://www.comp.nus.edu.sg/~cs3244/2010/e1.part1.html` to reach the entry form for this survey if you lose your browser window.
- Do remember that you will need to key in and submit your answers to the according assessment system as designated by your proctor or by exam central.

ANSWERS VERSION 1

1. (MRQ with 4 options; 2 marks) Tribes of ML. Mark which tribes' algorithms use some form of randomization as taught in lecture.

   (a) Analogizer.
   (b) Connectionist.
   (c) Symbolist.
   (d) Bayesian.

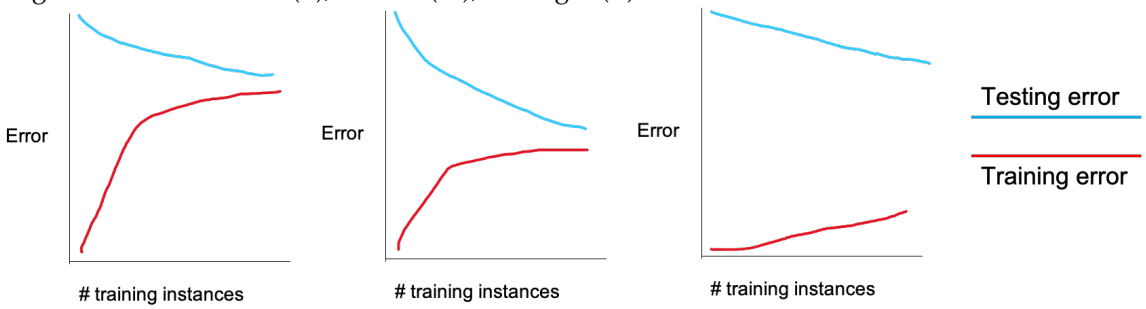
   Correct answers: (b)

2. (MRQ with 4 options; 3 marks) Which of the following statements is true regarding KNN?

   (a) KNN can be applied on regression tasks.
   (b) Using smaller k value leads to a smoother classification boundary.
   (c) As the k value increases, the classification accuracy will be higher.
   (d) The time complexity for training KNN is $O(1)$ and $O(n)$ for testing, where n is the number of test examples.

   Correct answers: (a), (d)

3. (MRQ with 5 options; 2 marks) Mark all correct associations with the diagrams below, where the diagrams are labeled left (L), middle (M), and right (R).



   (a) L — high bias
   (b) R — high variance
   (c) L — high variance
   (d) M — ideal
   (e) R — high bias

   Correct answers: (a), (b), (d)

[Questions 4–7] (MCQ; 1 mark each) Mark (a) for true and (b) for false for each of the following statements.

4. Supervised learning requires labeled data while unsupervised learning does not.

   Correct answers: (a)

5. KNN and K-means are both examples of unsupervised learning algorithms.

   Correct answers: (b)

6. The goal of supervised learning is to learn a target function that approximates the hypothesis function.

   Correct answers: (b)

7. SVM can penalise instances that are correctly classified.

   Correct answers: (a)

8. (MCQ; 2 marks) For which of the following tasks should you **not** use machine learning?

   (a) Create an agent that can fly a plane in a flight simulator.
   (b) Sort a list of random numbers.
   (c) Determine whether a student's transcript fulfills graduation requirements.
   (d) Predict the next number in a game of roulette after observing many rounds.
   (e) Predict the winner of a soccer tournament based on player statistics.
   (f) Predict the score of a student's machine learning exam given forum participation.

   Correct answers: (b), (c), (d)

9. (MCQ; 2 marks) KNN may overfit datasets of high dimensionality. Which option would you choose to handle this problem?

   (a) Non-linear mapping to fewer dimensions (dimensionality reduction).
   (b) Using a smaller K.
   (c) Both non-linear mapping to fewer dimensions and using a smaller K.
   (d) Neither of them.

   Correct answers: (a)

10. [This question not admissible for AY21/22S1] (MCQ; 2 marks) Suppose a perceptron with two input features does not have a bias weight. Mark the implication of this design choice.

    (a) It will require, on average, 1.5 times as many training examples compared to if it had a bias weight.
    (b) It will function as a sigmoid unit.
    (c) Its decision boundary will pass through the origin.
    (d) It will require, on average, 0.67 times as many training examples compared to if it had a bias weight.

    Correct answers: (c)

11. (MCQ; 2 marks) In which of the following tasks should you **not** use unsupervised machine learning techniques?

    (a) Identify irregular packet traffic on a network.
    (b) Group customer data.
    (c) Predict lottery results.
    (d) Help a farmer sort produce based on camera footage.
    (e) We can use it for all of the above.

    Correct answers: (c)

12. (MRQ with 5 options; 3 marks) SVM Knowledge. Mark the correct statement(s).

    (a) The hard SVM, like the perceptron, requires the dataset to be linearly separable.
    (b) SVMs achieve can find non-linear classification boundaries using a kernel function to map a $\mathcal{Z}$ space.
    (c) The margin is defined as the distance between the closest positive and closest negative point(s) in a training dataset to the optimal separative hyperplane.
    (d) The SVM penalises misclassifications more when they are further away from the margin boundary of their correct class.
    (e) [This option not admissible for AY21/22S1] The margin is defined to have unit (1) distance for convenience. Other values would also work.

    Correct answers: (a), (b), (c), (d), (e)

13. [This question not admissible for AY21/22S1] (MRQ with 5 options; 3 marks) By running the Candidate-Elimination algorithm on a task, a ver-sion space is produced with a specific S boundary sets and a general boundary set G.
Given a new input instance x, which of the following statements are correct?

(a) If x satisfies none of the members of G, then x satisfies none of the hypotheses in the version space.
(b) If x satisfies every member of S, then x satisfies every hypothesis in the version space.
(c) If x satisfies every member of G, then x satisfies every hypothesis in the version space.
(d) If x satisfies none of the members of S, then x satisfies none of the hypotheses in the version space.
(e) If x satisfies both S and G, there is only one hypothesis remaining in the version space.

Correct answers: (a), (b)

This marks the end of this part of the exam.
These is no additional material beyond this point.