

Name:

# Practice Midterm Exam Questions

CSE 142 – Machine Learning

November 1, 2021

14:40-15:40pm

**With camera on all the time.** All the teaching staffs will be watching and check the roster from times to times.

No phones. No earphones. No internet search. No keyboard typing.

Write answers on a white paper using your pen.

Be sure to read each question carefully and provide all the information requested. **If the question asks you to explain, do so!**

You will be given extra 5 minutes to upload your answers. Exams must be turned in by 15:45pm sharp.

Note that we provide as many questions as we can for you to practice, but the midterm exam will have less questions to fit in 1-hour duration.

Good luck!

### Sample Midterm Questions

1. [4 points] What is machine learning? Briefly describe the three key components of a learning problem?
2. [2 points] Give an example of an unsupervised discrete learning problem.
3. [3 points] How are training, validation, and testing data sets typically used in developing a machine learning algorithm or application?
4. [4 points] Describe inductive learning and deductive learning. Which do we focus on in machine learning and why?
5. [4 points] Briefly discuss the relationship between overfitting and generalization in machine learning. In general, what is likely to have lower error on the training data, a linear model or a higher-order polynomial model (and why)? How about on the testing data?
6. [3 points] What is the intrinsic dimensionality of a set of data?

7. [3 points] Give an example of a predictive machine learning task and an example of a descriptive machine learning task.
8. [6 points] A test for a new, deadly strain of anthrax (that has no symptoms) is known to be 99.9% accurate. The chances of any random person having this strain are one in a million. You get tested for anthrax during a routine medical exam, and your test comes back positive. If  $A$  is the variable that describes whether you have anthrax (true) or not (false), and  $T$  is the variable that describes the output of your anthrax test (true if you test positive, false if you test negative), what is the relative likelihood that you have anthrax? Use Bayes' Rule.
9. [4 points] In the basic binary linear classifier (with the linear discriminant function midway between the class centroids), the centroid of our positive class is at  $(2, 1, 4)$  and the centroid of our negative class is at  $(4, 4, 6)$ . The decision boundary is a plane defined by the vector  $\mathbf{w}$  and the threshold  $t$ . If we now add a new positive sample at  $(0, -2, 2)$  to the training data and recompute, how (qualitatively) will this affect the placement of the decision plane?

10. [3 points] You are given the probability tables for  $P(\text{data} \mid \text{hypotheses})$  and  $P(\text{hypotheses})$ . You need to choose the best model (hypothesis) from this data. What kind of decision rule is this?
11. [4 points] Why is feature selection often performed in a machine learning problem before learning the model?
12. [3 points] In testing the binary classification model you learned from the training data, you get 70 out of 100 instances correct. 40 of those correctly estimated the positive class (the concept), and the rest correctly estimated the negative class. There were a total of 60 positive examples in the test set. What is the false positive rate? The false negative rate? The precision? The recall?
13. [4 points] In learning a classifier, you use a loss function in weighing the effects of various training data instances. If the classification margin of an instance is very high, what should the loss function for that instance be (qualitatively, not a specific value)? If the margin of an instance is very low (negative), what should the loss function for that instance be? If an instance is badly misclassified, what should the loss function for that instance be?
14. You need to estimate prior probabilities for your 5-class classifier, and you have { 10, 8, 14, 9, and 5 } samples from your classes. What are your estimated prior probabilities, using Laplace correction?

15. Describe how a loss function (for classification or regression) can be made robust to outliers.
16. In concept learning, what is the difference between a *hypothesis* and a *concept*?
17. Give an example of a hypothesis that is not learnable by the conjunctive hypothesis space representation.
18. In our conjunctive hypothesis space learning, we generally seek the least general generalization. Compared with a more general generalization, what effect does choosing a less general hypothesis have on our false positive rate?
19. What hypothesis will guarantee a *consistent* concept for any problem?

20. In a decision tree learning routine, a particular node has 8 examples of the positive class and 2 examples of the negative class. What is the impurity measure of that node?
21. A decision tree for concept  $c$  has five leaves with the following training examples in each leaf:
- L1: (5 pos, 2 neg)
  - L2: (6 pos, 1 neg)
  - L3: (3 pos, 4 neg)
  - L4: (0 pos, 3 neg)
  - L5: (1 pos, 2 neg)

Using Laplace correction, give the ranking order of the leaves (from highest ranked to lowest).

22. In a univariate linear regression problem, what is the geometric interpretation of the regression coefficient?