CSE 417T: Exam 2 Practice Questions

1 Written-response questions

1. Consider the following data set: the class variable is whether or not a car gets good gas mileage GoodMPG? (Yes or No), and the features are Cylinders (either 4 or 8), Displacement (High, Medium, or Low) and Horsepower (High, Medium, or Low).

Cylinders	Displacement	Horsepower	GoodMPG?
4	Low	Low	No
4	Low	High	Yes
4	Medium	High	No
8	High	Low	Yes
8	Medium	Medium	Yes
8	High	Medium	No

Given this data set, what feature would the ID3 algorithm split on first? Hint: you should be able figure this out without explicitly computing any of the information gains.

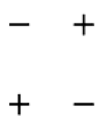
Will the ID3 algorithm learn an optimal (fewest number of splits) decision tree that minimizes the in-sample error on this data set? Explain why.

2. Suppose you are running AdaBoost using depth-0 decision trees as the weak learners (depth-0 decision trees simply predict whichever class has the most total weight across all training data points). Assume 80% of the training data points are positive examples and 20% of them are negative examples. After running one round of AdaBoost, what is the total weight of all the positive examples?

Given your response to the previous question, what weak leaner would you learn in the second iteration? Based on these findings, do you think that depth-0 decision trees are a good choice of weak learner for the AdaBoost algorithm? Why or why not?

3. Assume you are given a decision tree that classifies binary vectors with 100 features. Can you specify a (1-)nearest neighbor model (by constructing a training data set) that would predict exactly the same classification of all possible data points as the decision tree? Explain how, or provide a counter-example.

4. Consider the following data set:

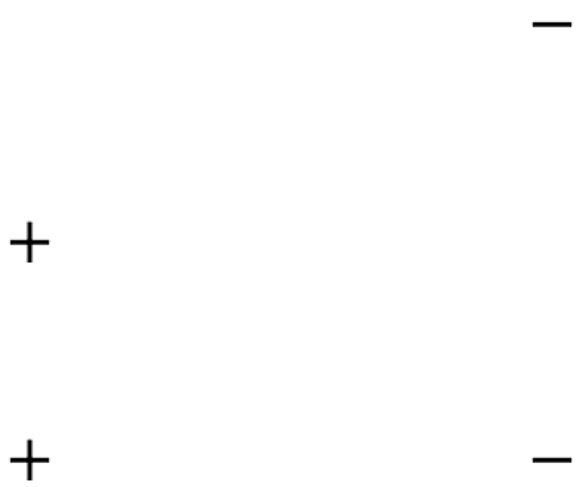


Which of the following classifiers will achieve zero training error on this data set? Select all that apply.

- ◯ SVMs
- ◯ Depth-1 decision trees
- ◯ A 3-NN classifier
- ◯ Neural networks with no hidden layers

For each of the options you did not select, propose a slight modification that would allow that model to achieve zero training error on this data set?

5. Consider the following data set:



Draw the decision boundary you would learn using a hard-margin SVM on this data set.

A non-essential support vector is a support vectors that if removed, would not result in a different hard-margin SVM decision boundary. An essential support vector is a support vectors that if removed, would change the hard-margin SVM decision boundary. Indicate in the figure above all non-essential support vectors.

Next, add a positive point that a) changes the hard-margin SVM decision boundary and b) increases the number of essential support vectors and decreases the number of non-essential support vectors. How many essential support vectors and non-essential support vectors are there in the new data set?

2 Multiple choice questions

6. Given some domain $\mathcal{X}$, let $H_2(\mathcal{X})$ be the set of all depth-2 decision trees and $H_\infty(\mathcal{X})$ be the set of all possible decision trees. Which of the following statements is true?

   ◯ $\text{Bias}(H_2(\mathcal{X})) < \text{Bias}(H_\infty(\mathcal{X}))$

   ◯ $\text{Bias}(H_2(\mathcal{X})) = \text{Bias}(H_\infty(\mathcal{X}))$

   ◯ $\text{Variance}(H_2(\mathcal{X})) = \text{Variance}(H_\infty(\mathcal{X}))$

   ◯ $\text{Variance}(H_2(\mathcal{X})) < \text{Variance}(H_\infty(\mathcal{X}))$

7. Given n training data points, what is the maximum depth of a tree returned by the ID3 algorithm?

   ◯ 2

   ◯ $\log_2(n)$

   ◯ n

   ◯ 2^n

8. For AdaBoost, recall that the update rule for the weight on the i^{th} data point is

$$\omega_i^{(t)} = \frac{\omega_i^{(t-1)} e^{-\alpha_t y_i h_t(\vec{x}_i)}}{Z_t}$$

   What is the weighted training error of h_t using the updated (and normalized) weights $\omega_i^{(t)}$?

   ◯ 0

   ◯ ϵ_t

   ◯ 0.5

   ◯ 1

9. What is the relationship between the importances of weak learners learned in consecutive iterations of AdaBoost?

   ◯ $\alpha_{t-1} < \alpha_t$

   ◯ $\alpha_{t-1} = \alpha_t$

   ◯ $\alpha_{t-1} > \alpha_t$

   ◯ None of the above

10. What is one reason why you might prefer lower values of k over higher values of k in a k-NN model for a particular problem?

- ◯ Lower values of k do not have an associated training time
- ◯ Lower values of k tend to give more complex decision boundaries
- ◯ Predictions with lower values of k can be computed faster
- ◯ Predictions with lower values of k tend to have higher accuracy

11. Recall that one commonly used weight function in the RBF model is $\phi(z) = e^{-\frac{z^2}{2}}$. Which of the following would be a reasonable alternative to this weight function?

- ◯ $\phi(z) = z$
- ◯ $\phi(z) = z^{-2}$
- ◯ $\phi(z) = \log\left(\frac{z}{2}\right)$
- ◯ $\phi(z) = e^{\frac{z^2}{2}}$

12. What kinds of functions can be implemented by two-layer neural networks that use the activation functions $\theta(s) = Cs$ for some constant C?

- ◯ Constant functions
- ◯ Linear functions
- ◯ Continuous functions
- ◯ Smooth functions

13. What kind of features are detected by the following convolutional filter:

$$\begin{bmatrix} 0 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

- ◯ All edges
- ◯ Only vertical edges
- ◯ Only horizontal edges
- ◯ Both horizontal and vertical edges