CS 491 Midterm 1 Name _____

Short Problems (20 points)

1. Given the following categorical data $\{Noun, Verb\}$, how could you adapt the categorical data such that you could use K-means clustering on it? Give what K value you should use. (4 points)

2. If u and v are any two orthogonal unit vectors, then $||u + v||_2 = 1$. Orthogonal=Perpendicular. Unit=Length is 1. True or false. If true, prove it. If false give a counter example. (4 points)

3. If the training data is linearly separable, then the 3-nearest neighbors algorithm will always have 100% accuracy on the training set. True or False. Explain your answer. (4 points)

4. The decision tree classifier has 100% accuracy on the training set (namely, the data is noise-free). Will a linear classifier have the same accuracy (100%) on the training set? Explain your answer. (4 points)

5. Compute the gradient of the following function at $(1, 1, 1)$: $f(x, y, z) = 2x^2+3y^3+4x^4+xyz+3x^2y+4y^2z$ (4 points)

Decision Trees (20 points)

6. Use the following data for the 2D XOR problem.

Sample	x_1	x_2	Label
s_1	-1	-1	0
s_2	-1	1	1
s_3	1	-1	1
s_4	1	1	0

(a) Does it make sense to generate a depth-1 decision tree for 2D XOR? Why or why not? (5 points)

(b) Generate the best depth-2 decision tree for the 2D XOR problem. (10 points)

(c) How would your decision tree from part (b) classify the following sample? (5 points)

Sample	x_1	x_2	x_3	Label
s_t	-1	-1	1	0

Optimization (15 points)

7. Recall our Regularized Optimization problem to find a linear separator given non-linearly separable data. We are trying to find the w and b that minimize the following objective function.

$$\underset{w,b}{\text{minimize}} \quad 1[y(w \bullet x + b) \leq 0] + \lambda R(w, b)$$

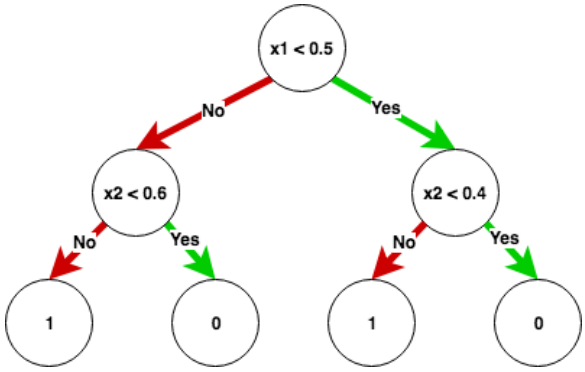
   (a) What are the two terms in the above equation doing? Explain them individually. That is, explain what $1[y(w \bullet x + b) \leq 0]$ is doing and explain what $\lambda R(w, b)$ is doing. (10 points)

   (b) Assuming R does the "right thing," what value(s) of λ will lead to overfitting? What value(s) will lead to underfitting? (5 points)

491

Decision Boundaries (15 points)

8. Consider the following Decision Tree.



(a) Draw the decision boundary for this tree, labeling all areas with the correct class. (10 points)

(b) Suppose we perform 1-nearest neighbor classification, instead of using the decision tree given above. The training data has four samples from each class. Is it possible that we obtain the same decision boundaries for the 1-NN classifier that we got for the decision tree in part a? If yes, give an example of the location that the points could have. If no, explain why. (5 points)

491

Linear Classifier (20 points)

9. Suppose we have the following training data.

Sample	x_1	x_2	Label
s_1	0	0	-1
s_2	1	0	-1
s_3	0	1	-1
s_4	2	0	1
s_5	1	1	1
s_6	0	2	1

(a) Give the weights w_1, w_2, and b for a neuron that perfectly classifies the training data. (10 points)

(b) Draw the decision boundary for your classifier. (5 points)

(c) How would your classifier classify the following test sample? $s_t = (1.5, 1, 1)$ (5 points)

491

K-Means (10 points)

10. Given the three following sets of data (i, ii, and iii). Assume you want to cluster each set of data into two clusters. Explain, and draw, what would likely happen with K-Means (K=2) in each case and why.

i

ii

iii