CS 491 Midterm 1 Name ____________________________

Short Problems (20 points)

1. For every training data set $(x_1, y_1), \ldots, (x_n, y_n)$, the training error of the 1-nearest neighbor classifier is always zero. You may assume that each x_i is unique. True or False. If true, explain. If false, give a counter example. (4 points)

2. When building a decision tree, we never ask about the same feature twice. True or False. Circle one and explain. (4 points)

3. I have trained a perceptron on a set of data and achieved 100% accuracy on the training data. Will a decision tree necessarily achieve the same accuracy? Explain your answer. (4 points)

4. K-Means always produces a linear decision boundary. True or False. If true, explain. If false, give a counter example. (4 points)

5. What purpose do surrogate loss functions serve in gradient descent? (4 points)

491

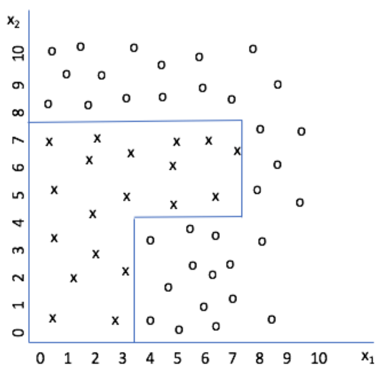
Decision Trees (15 points)

6. Use the following data.

Sample	x_1	x_2	Label
s_1	2	2	0
s_2	2	1	1
s_3	1	2	1
s_4	1	1	0

(a) Using entropy and information gain, at what depth would your algorithm stop building the decision tree? (5 points)

(b) Given the following decision boundary, draw a decision tree that would produce such a boundary. (10 points)



491

Optimization (20 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) Let us use the following loss function: $L(y, \hat{y}) = (y - \hat{y})^2$ and the following regularization: $R = ||w||^2$. Find $\nabla_w L$ and $\frac{\partial L}{\partial b}$

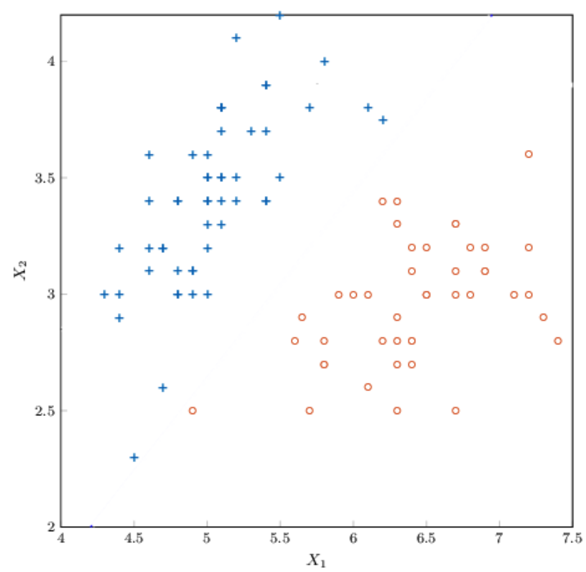
   (b) Explain how $R(w, b) = \sum_i w_i \neq 0$ enforces a "simpler" solution? (5 points)

   (c) What purpose does λ serve in this optimization problem? (5 points)

491

Decision Boundaries (15 points)

8. Consider the following data.



(a) Explain what the decision boundary looks like for K-NN with K=N? There are 41 + and 38 *o*. (5 points)

(b) Would a K-NN with K=3 perfectly classify the training data? Explain. (5 points)

(c) Could a depth-2 decision tree perfectly classify the above data? Draw the best (with respect to accuracy) depth-2 decision tree. (5 points)

491

Linear Classifier (20 points)

9. Suppose we have the following training data.

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

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

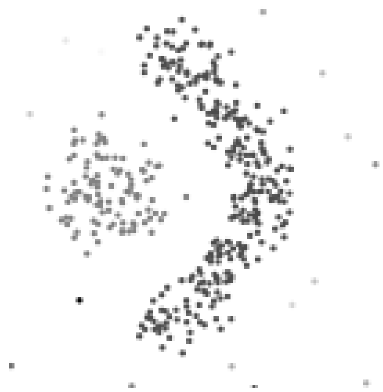
(b) Would you get the same weights and bias if you iterated from s_4 up through s_1? (5 points)

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

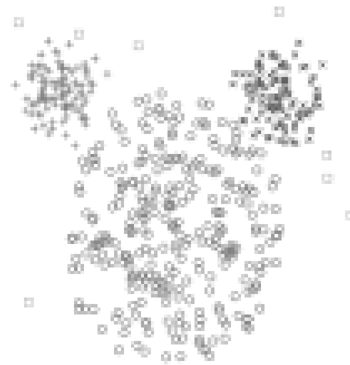
491

K-Means (10 points)

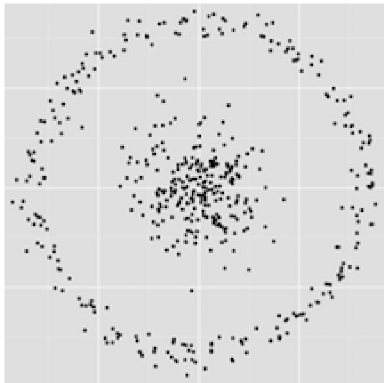
10. Given the three following sets of data (i, ii, and iii). Assume you want to cluster each set of data into clusters. Explain, and draw, what would likely happen with K-Means in each case and why.



(i) K=2



(ii) K=3



(iii) K=2

491

$$H() = -\sum_{c} p(c) log_2 p(c)$$

$$IG = H() - \sum_{t} p(t) H(t)$$