CS 691 Midterm 1 Name _____

Short Problems (20 points)

1. We can only have a tie in the case of KNN when K is even. True or false. If true explain. If false, give a counter example. (4 points)

2. We are trying to apply machine learning to a particular problem. We have a labeled dataset consisting of 10 million samples, each containing 10 real-valued features. What would be the best algorithm/model for this problem? What would be the worst? 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. Gradient descent is guaranteed to give you a global minimum. True or False. If true, explain. If false, give a counter example. (4 points)

691

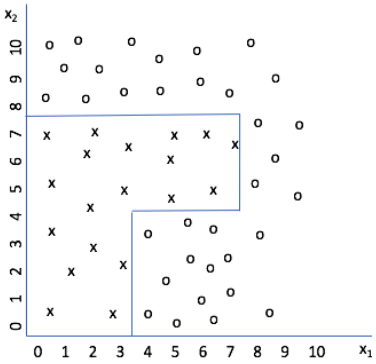
Decision Trees (15 points)

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)



691

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.

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

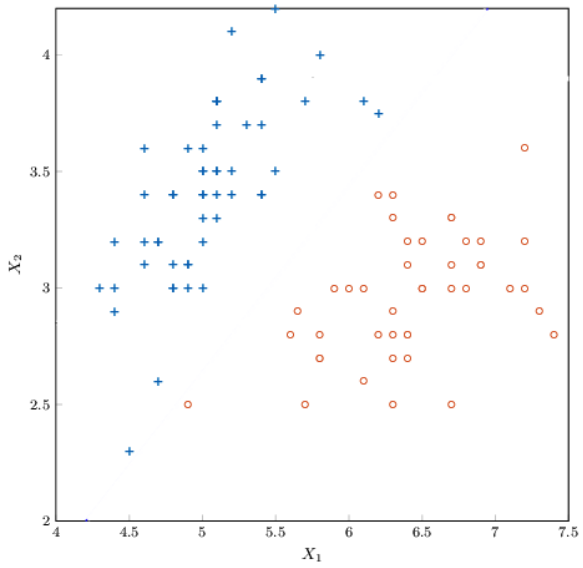
(a) Let us use the following loss function: $L(y, \hat{y}) = (y - \hat{y})^2$ and the following regularization: $R = ||w||^2$. Compute the $\nabla_w L$ and $\frac{\partial L}{\partial b}$. (10 points)

(b) Assume we have two nearby samples: $s_1 = [x_1, x_2, ..., x_d]$ and $s_2 = [x_1+\epsilon, x_2, ..., x_d]$. s_2 is the same as s_1 except that the first feature is off by a small number ϵ. Show how adding $R(w, b) = ||w||^2$ to our optimization problem helps to ensure that neighboring samples have similar predictions. (10 points)

691

Decision Boundaries (15 points)

8. Consider the following data.



(a) Could K-NN perfectly separate the data? If so, for what values of K? If not, explain. (5 points)

(b) On the figure, draw the decision boundary for a perceptron. Could you get a different boundary with a perceptron? (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)

691

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)

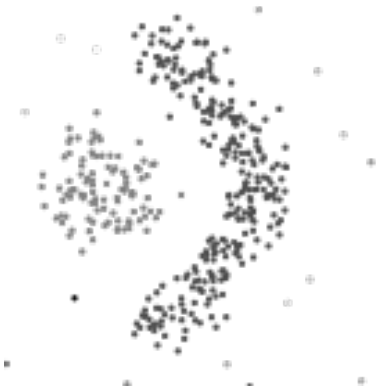
(b) Would you get the same weights and bias if you iterated from s_4 up through s_1? (5 points)

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

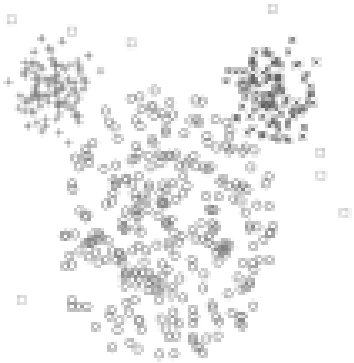
691

K-Means (10 points)

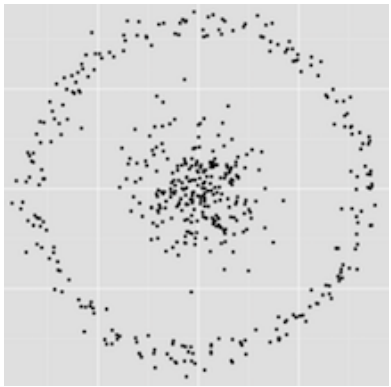
10. Given the three following sets of data (i, ii, and iii). Assume you want to cluster each set of data into clusters. Explain, and draw, what would likely happen with K-Means in each case and why.



(i) K=2



(ii) K=3



(iii) K=2

691

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

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