CS 422/622 Exam 1 Name ____________________________

Short Answer (10 Points)

1. (422) True/False: A Decision Tree is more powerful than a perceptron. That is, it can solve more complex problems. If false, briefly explain. If true, give an example.

1. (622) Given that my training data has f binary features and s samples, give an expression for the maximum depth a decision tree can have in terms of these two values.

2. Why is it that a different ordering of the input data may result in a different w and b output from the perceptron?

Short Answer (10 Points)

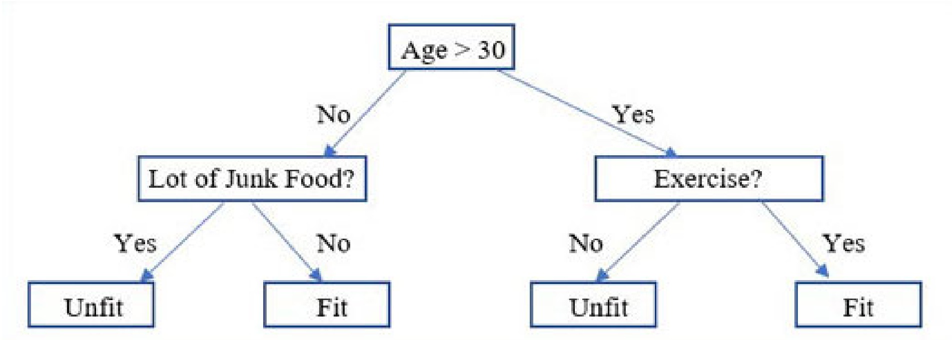
3. (422) True/False: KNN and K-Means behave similarly with noisy training samples. Briefly explain.

3. (622) True/False: Duplicate training samples (same features and label) have no effect on the output of K-Means. For example, if you run K-Means on a dataset with no duplicates you will get the same result as you would if you added 20% duplicates to that same dataset. If true, briefly explain. If false, give an example.

4. "I don't remember all the details of the adaboost algorithm, but I do remember that the final prediction is a weighed sum of the prediction of weak learners." Is this an example of poor recall or precision? Briefly explain.

Adaboost (10 Points)

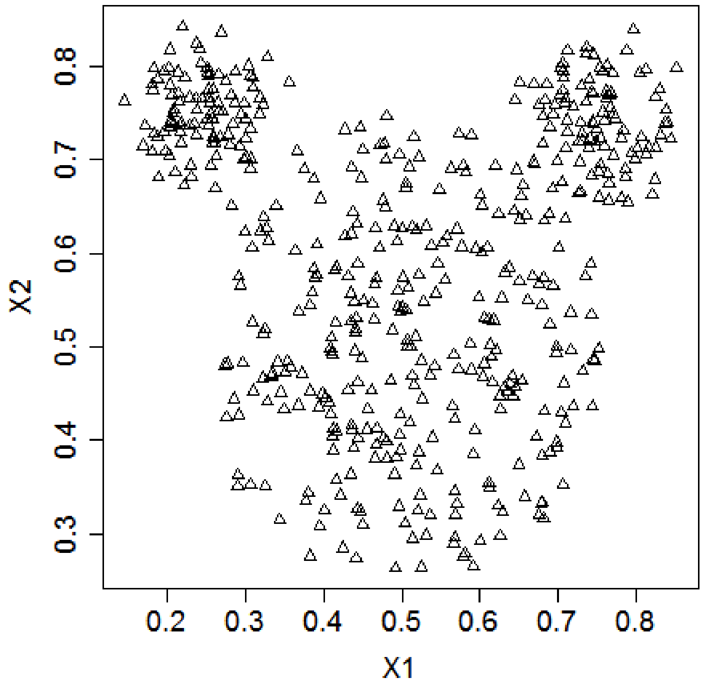
5. Assume you have the following decision tree and training data. Give the sample weights for the first and second round of Adaboost (the final two columns of the table) using this decision tree as your week learner.



Sample	Age	Junk Food	Exercise	Label	$d^{(0)}$	$d^{(1)}$
s_1	20	0	0	0		
s_2	21	0	1	1		
s_3	27	1	0	0		
s_4	28	1	1	0		
s_5	31	0	0	1		
s_6	33	0	1	1		
s_7	38	1	0	1		
s_8	40	1	1	0		

K-Means (10 Points)

6. How would K-means cluster the following data? Indicate the cluster centers and the rough clusters on the graph. Explain your choices briefly.



Linear Classifiers (20 Points)

7. Give the gradient descent update rules for the following regularized loss function: $L = e^{y\hat{y}} + \lambda|w|$.

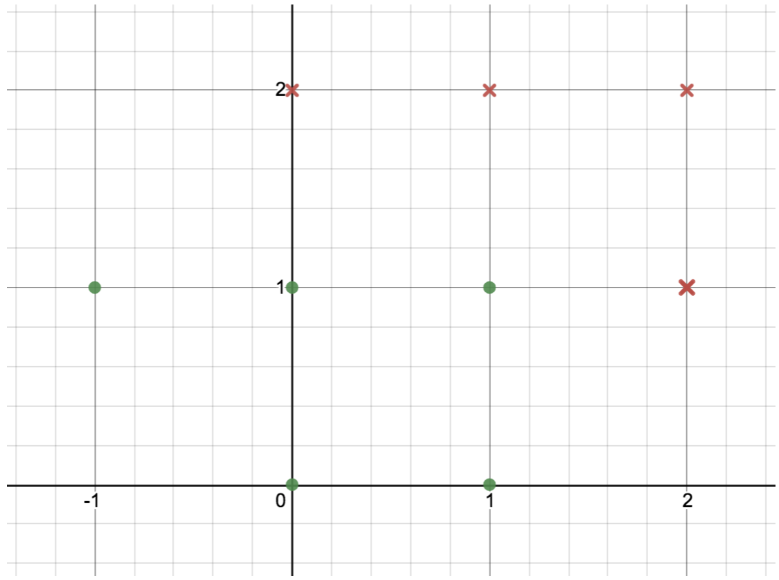
Decision Tree (20 Points)

8. Build a decision tree using the greedy information gain algorithm from class using the following training data. Given a tie between features x_i and x_j choose x_i such that $i < j$. That is, if there is a tie between x_2 and x_3, choose x_2.

Sample	x_1	x_2	x_3	y
s_1	0	0	0	0
s_2	0	0	1	1
s_3	0	1	0	0
s_4	0	1	1	0
s_5	1	0	0	1
s_6	1	0	1	1
s_7	1	1	0	1
s_8	1	1	1	0

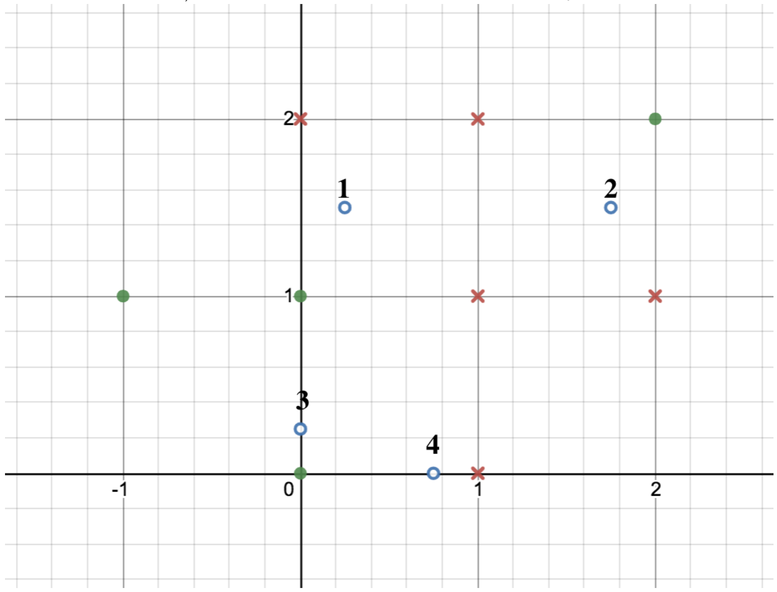
Perceptron (10 Points)

9. Give a w and b for a linear classifier that perfectly classifies the following data (x = negative, circle = positive). Show/explain how you came up with these values.



KNN (10 Points)

10. Using K=1, 3 and 5, classify the test data (open circles) using the training data (x = negative, filled circle = positive). Use the table below to record your final answers.



Sample	$K=1$	$K=3$	$K=5$
1			
2			
3			
4			

Equations

Entropy and Information Gain

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

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

p	$plog(p)$
$\frac{1}{8}$	-0.375
$\frac{1}{4}$	-0.5
$\frac{3}{8}$	-0.53
$\frac{1}{2}$	-0.5
$\frac{5}{8}$	-0.423
$\frac{3}{4}$	-0.311
$\frac{7}{8}$	-0.168
1	0

Adaboost

Algorithm 32 ADABOOST($\mathcal{W}$, $\mathcal{D}$, K)

$d^{(0)} \leftarrow \langle \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \rangle$ // Initial

for $k = 1 \dots K$ **do**

$\quad f^{(k)} \leftarrow \mathcal{W}(\mathcal{D}, d^{(k-1)})$

$\quad \hat{y}_n \leftarrow f^{(k)}(x_n), \forall n$

$\quad \hat{\epsilon}^{(k)} \leftarrow \sum_n d_n^{(k-1)} [y_n \neq \hat{y}_n]$

$\quad \alpha^{(k)} \leftarrow \frac{1}{2} \log \left(\frac{1 - \hat{\epsilon}^{(k)}}{\hat{\epsilon}^{(k)}} \right)$

$\quad d_n^{(k)} \leftarrow \frac{1}{Z} d_n^{(k-1)} \exp[-\alpha^{(k)} y_n \hat{y}_n], \forall n$

end for

return $f(\hat{x}) = \text{sgn} \left[\sum_k \alpha^{(k)} f^{(k)}(\hat{x}) \right]$

Perceptron

$$a = w \cdot x + b$$

$$w = w + xy$$

$$b = b + y$$