Name: Jamia Khana

Problem	Points	Score
1(a)	20	
1(b)	10	
2(a)	10	
2(b)	20	
2(c)	10	
3(a)	10	
3(b)	10	
3(c)	10	
Total	100	

Notes:

(1) The exam is closed books and notes.

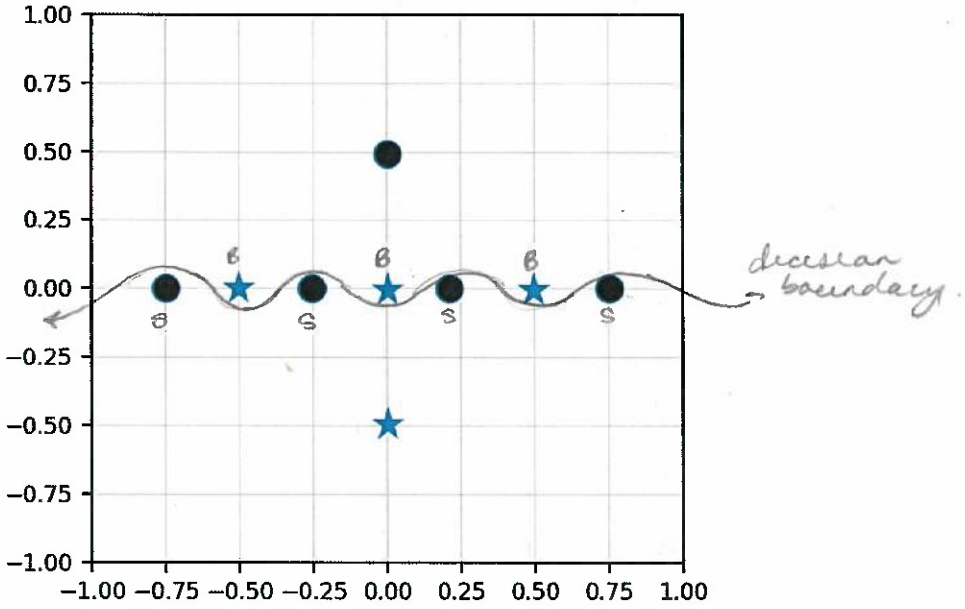
(2) Please clearly indicate your answer to the problem.

(3) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.

(4) If I can't read or follow your solution, it is wrong, and no partial credit will be awarded.

(40 pts) Problem No. 1: Given the data shown below:

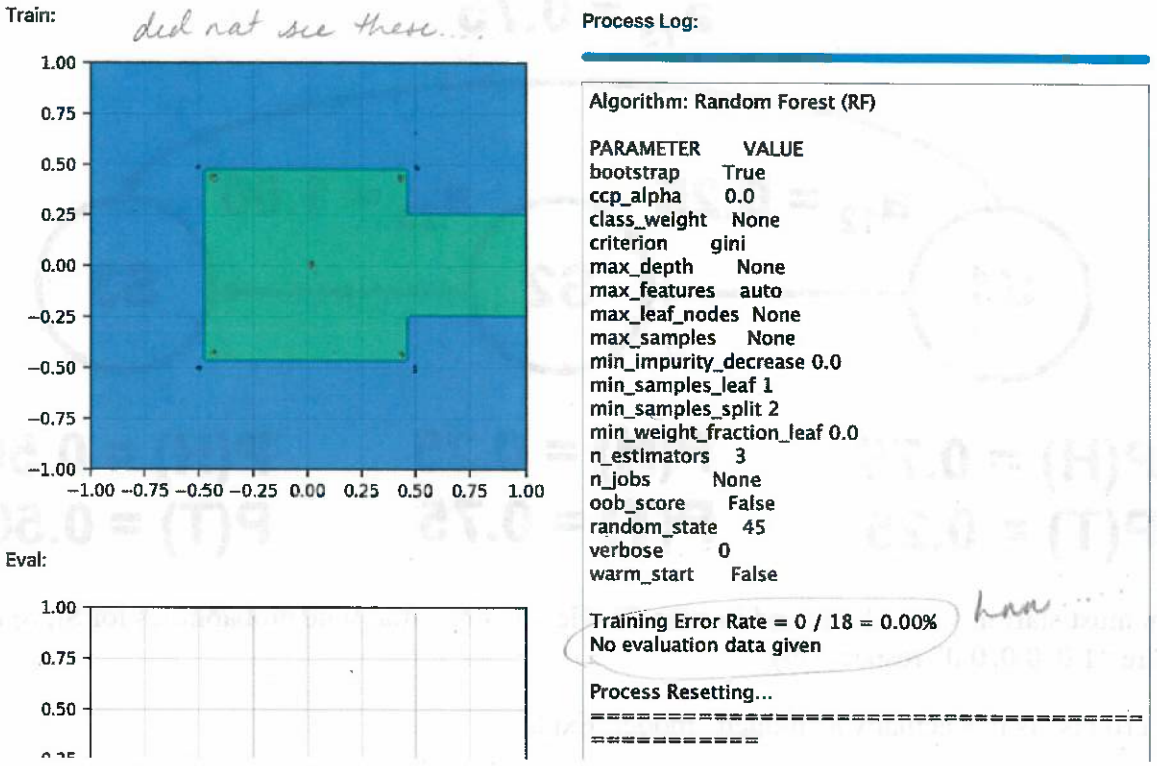
(a) (20 pts) Draw the decision surface you would obtain if you applied the k-nearest neighbor algorithm (KNN) with $K = 3$ to this data. Justify your result with a detailed explanation.



Justification:

When using KNN, we want to look at however many neighbors and take the majority vote given our value of K. For this question, we are given a K-value of 3 which means we will take the majority vote of 3 neighbors for any given data point. A KNN algorithm of K=3 would do incredibly poorly as seen above. The classes that each datapoint would be assigned, given K = 3, have been written by the data point where 'S' represents class star and 'B' represents black. In the middle row where y = 0, each data value will be assigned the class of its 2 neighbors as they are the closest data value and 2/3 will always be the majority vote for K = 3. Thus, the decision boundary for K = 3 will result in a line that looks similar to a sin wave that snakes in and out of the values and assigns them to the wrong class.

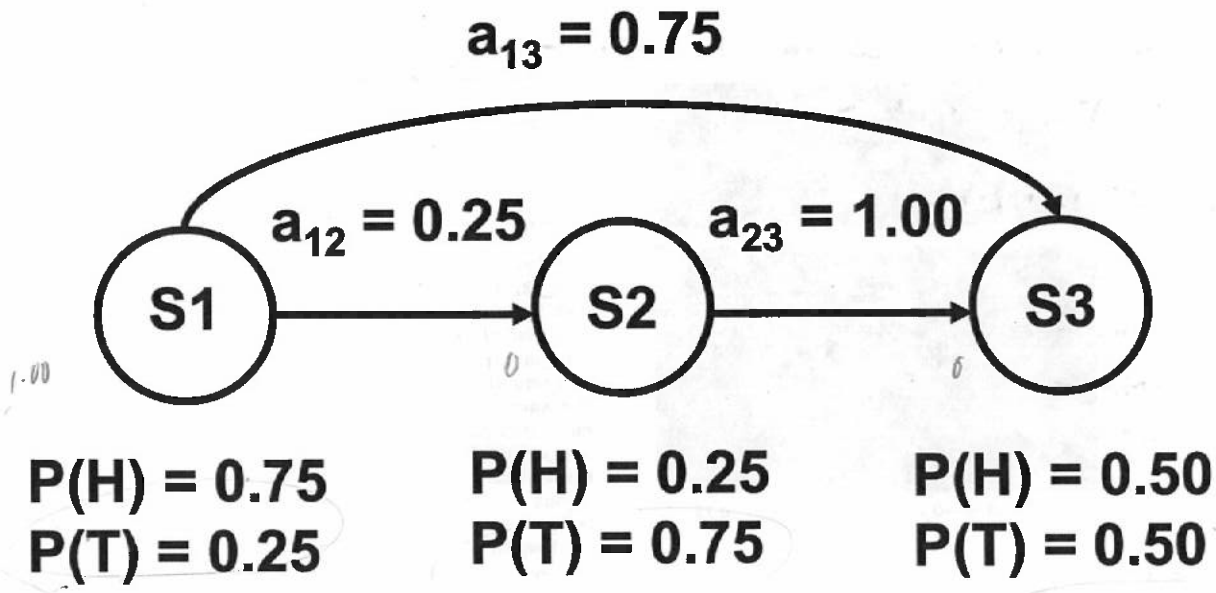
(b) (10 pts) I was playing around with IMLD before the exam and discovered a bug when comparing results with JMP. Consider the result shown below:



This was generated with Random Forests using 3 trees. Do you agree with IMLD? Justify your answer.

i do agree with the outputted graph w/ 3 nodes. i think that it is possible to produce such a graph with 3 trees and however many leaves per each. I also agree that it is possible to attain an error rate of 0% with random forest but this will lead to an inability to generalize on later test data (probably). Thus i do agree with the output from imld in terms of illustration and output, but i do not agree that this is good. i think that we would want to prune this in order to attain some error rate to ensure that we can generalize later test data. i did not see the orange values in the graph but it still do not think my answer changes.

Problem No. 2: Given the hidden Markov model shown below:



You must start in state S1 and end in state S3. Hence, the initial state probabilities for S_1, S_2 and S_3 are "1.0, 0.0, 0.0" respectively.

(a) (10 pts) Is this actually a "hidden" model. Explain.

i think that this model could be considered a "hidden" model but i also think that because you would be able to deduce the state of the model because of its simplicity that you could argue that it is not truly hidden. the idea behind a hidden markov model is that one is unable to determine the state of the model until an output is given. this idea is applicable to all states of the model. however, since we know that the model will always start in S1, then we always know the state of the model at initiation. and therefore know the probabilities of P(H) and P(T). With this in mind, i do not think you could say that this is truly a hidden markov model even though it is intended to be one.

(b) (20 pts) Assume you are given the training sequences: "HH", "HT", "TH", "TT", "HHH", "TTT", "HTHT", and "THTH". Reestimate the transition probabilities a_{12} and a_{13}.

So lets begen by placing our value in each of their respecter states.

S_1 HHTTHTHHT

S_2 HTTH

S_3 HTHTHT

(c) (10 pts) If you were to use this model to randomly generate data, what is the average duration of the sequences produced?

we know that from state 1, there is a $p = 0.75$ chance to transition to state 3 (terminal) and 0.25 to transition to state 2. To be honest, im not entirely sure that i fully grasp HMM however i am going to try. given these probabilities we know that there is likely to be a sequence of only 2 values 75% of the time ($S_1 \rightarrow S_3$). the rest of the time we will see more (lets say 3) ($S_1 \rightarrow S_2 \rightarrow S_3$). Thus, id like to say that the average sequence of values (duration) will be $2(0.75) + 3(0.25) \Rightarrow 1.50 + \sim 0.75 \approx 2.25$ is likely to be the averaged time duration or we may just say 2.

Problem No. 3: A discrete random variable, X, has a probability mass function (pmf):

$$p_k = \begin{Bmatrix} 1/3 & 0 \le k < 2 \\ 2/3 & 2 \le k < 4 \end{Bmatrix}.$$

0.33
0.66

A similar random variable, Y, has a probability mass function:

$$p_k = \begin{Bmatrix} 1/4 & 0 \le k < 2 \\ 1/4 & 2 \le k < 4 \end{Bmatrix}.$$

entropy is H.

Equations you might find useful for this problem include:

$$H(X) = -\sum_{x \in X} p(x) \log p(x)$$

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$$

yowza!

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x|y)}{p(x)}$$

$$I(X;Y) = H(X) - H(X|Y)$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

(a) (10 pts) Compute the entropy of X and Y. Explain why your answers makes sense.

$H(x) = -\sum p(x) \log(p(x))$

$H(x) = -(0.33 \log(0.33) + 0.66 \log(0.66))$

you definitely want to use the $H(X)$ & $H(x,y)$ equations.

$H(x) = -\sum p(x) \log p(x)$
$H(y) = -\sum\sum p(x,y) \log p(x,y)$

} once you calculate these two you have entropy x and entropy between x, y

(b) (10 pts) Assume the joint distribution between X and Y is a uniform distribution: $p(x, y) = 1/16$. Compute the mutual information. Justify your answer.

(c) (10 pts) Suggest a shape for the joint distribution that would increase the mutual information. Justify your answer.