CSE 485 Final

Olcay Taner YILDIZ

I. Question (4 points)

Define conditional probability and write down Bayes theorem.

II. Question (10 points)

a) In the Cherbonyl nuclear power station, there is an alarm which sense when a temperature gauge exceeds a given threshold. The gauge measures the core temperature. Let A = "alarm sounds" (yes, no); G = "gauge reads" (normal, high); T = "core temp" (normal, high); FA = "alarm is faulty" (yes, no); FG = "gauge is faulty" (yes, no). Draw a Bayesian network for this domain.
b) Suppose the gauge give incorrect temperature x% of the time when it is working, but y% of the time when it is faulty. Give the conditional probability table of G.
c) Suppose the alarm works unless it is faulty, in which case it never goes off. Give the conditional probability table of A.
d) Suppose we add a second temperature gauge H, and connect the alarm so it goes off when either gauge reads High. Update Bayesian network in a) accordingly.

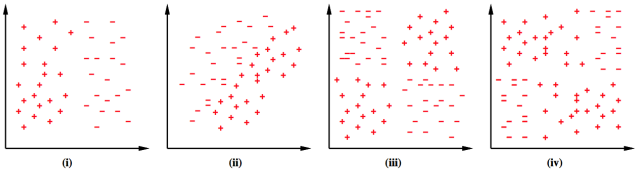
III. Question (8 points)

Consider the following data, each consisting of three inputs and a class: (111, 1) (110, 1) (011, 1) (010, 0) (000, 0).

a) Draw a decision tree consistent with the data.
b) Draw a linear perceptron consistent with the data.

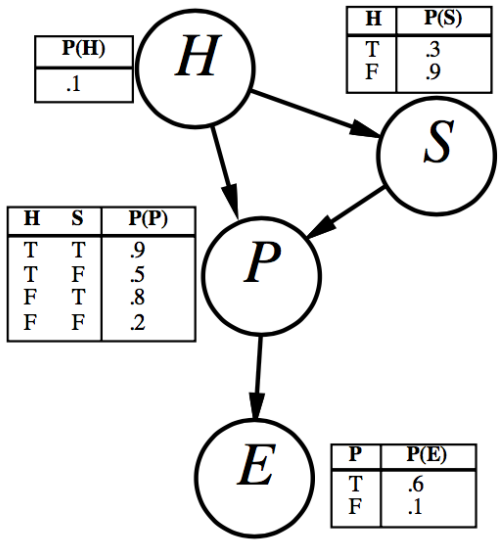
IV. Question (9 points)

Consider the following datasets.



a) Which of the above can be correctly classified by a linear perceptron?
b) Which of the above can be correctly classified by a low depth decision tree?
c) Which of the above can be correctly classified by a 3-NN?

V. Question (8 points)



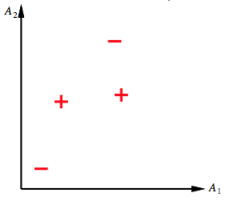
Consider the above Bayesian Network, where H = Honest, S = Slick, P = Popular, E = Elected.

a) Calculate the value of P(h = T, s = T, p = F, e = F).
b) Calculate the probability that someone is elected given that they are honest.

VI. Question (6 points)

Suppose we are given the following dataset.

A_1	A_2	Y
3	3	*false*
6	13	*true*
15	14	*true*
14	22	*false*



a) Draw a decision tree that classifies the examples correctly.
b) Calculate the entropy of your root node test.

VII. Question (8 points)

a) Suppose you train a classifier and test it on a held-out validation set. It gets 80% classification accuracy on the training set and 20% accuracy on

the validation set. From what problem is your model most likely suffering? Underfitting, overfitting? Explain.

b) Suppose you train a classifier and test it on a held-out validation set. It gets 30% classification accuracy on the training set and 30% accuracy on the validation set. From what problem is your model most likely suffering? Underfitting, overfitting? Explain.

VIII. Question (8 points)

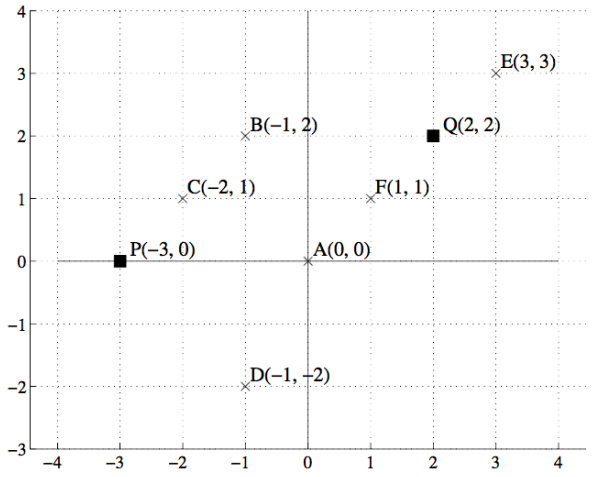
Consider the geometric distribution, which has $P(X = k) = (1 - \theta)^{k-1}\theta$. Assume in our training data X took on the values 4, 2, 7, and 9.

a) Write an expression for the log likelihood of the data as a function of the parameter θ. Remember, the likelihood is calculated by multiplying the likelihoods of the data points. Log likelihood is log of likelihood.

b) Take the derivative of the expression in part a) solve for θ and calculate the maximum likelihood estimate.

IX. Question (12 points)

Suppose we are given the following dataset. In this question, we will do k-means clustering to cluster the points A, B ... F into 2 clusters. The current cluster centers are P and Q.



a) Select all points that get assigned to the cluster with center at P.

b) What does cluster center P get updated to?

c) Suppose the new distance function is $d(A, B) = |A_x - B_x| + |A_y - B_y|$. Select all points that get assigned to the cluster with center at P.

d) What does cluster center P get updated to w.r.t. new distance function?

X. Question (8 points)

Consider the following two-class dataset. Your classifier considers the number of 7-letter (f7) and 8-letter words (f8) in an essay and then assigns a grade, either A or F, based on those two numbers. You have four graded essays to learn from. Suppose you run a linear perceptron, and if the score from your classifier is greater than 0, it gives an A, if it is 0 or lower, it gives an F.

$BIAS$	f_7	f_8	grade
1	2	1	A (+)
1	0	2	F (-)
1	1	2	A (+)
1	1	0	F (-)

Given your initial weight vector as (1, 0, 0), what will be your weight vector after having seen the first training example and after having seen the second training example.

XI. Question (9 points)

A 4200 nt long DNA sequence is used as a training set for parameter estimation of the DNA statistical model. The observed counts of sixteen dinucleotides, N_{XY}, are as follows.

	T	C	A	G
T	510	380	210	190
C	240	170	360	230
A	370	200	220	210
G	190	170	220	220

Find the maximum likelihood estimates of the transition probabilities P_{TT}, P_{AG}, P_{CT} of the first order Markov model of the DNA sequence.

XII. Question (10 points)

Given the following error rates for a four feature problem, what will be the result of the following attribute subset selection procedures?

- Stepwise forward selection
- Stepwise backward elimination

Error Rates: (F_1), 0.23; (F_2), 0.25; (F_3), 0.36; (F_4), 0.24; (F_1, F_2), 0.18; (F_1, F_3), 0.19; (F_1, F_4), 0.20; (F_2, F_3), 0.16; (F_2, F_4), 0.29; (F_3, F_4), 0.22; (F_1, F_2, F_3), 0.39; (F_1, F_2, F_4), 0.45; (F_1, F_3, F_4), 0.24; (F_2, F_3, F_4), 0.36; (F_1, F_2, F_3, F_4), 0.59.