CSE 560 Final

Olcay Taner YILDIZ

I. QUESTION (18 POINTS)

1) Consider a belief network where there are four nodes A, B, C, D. There are three arcs one from A to B, one from B to C and one from B to D. Which of the following statements are implied by the network structure?
   - $P(B|A, C, D) = P(B|A)$
   - $P(D|B) = P(D|B, C)$
   - $P(B|A) \neq P(B)$
2) Suppose we have evidence at B. Use Bayes' rule to solve the query $P(A|B)$ in terms of probabilities directly available in the network. (You may assume the usual normalization constant α.)
3) Suppose instead that we have evidence at C and D. Use Bayes' rule followed by conditioning on B solve the query $P(A|C, D)$ in terms of probabilities directly available in the network. [Given an expression to $P(X|Z)$, conditioning on Y gives $P(X|Z) = \sum_y P(X|Y = y, Z)P(Y = y|Z)$.]

II. QUESTION (28 POINTS)

1) Consider a single Boolean random variable Y (The 'classification'). Let the prior probability P(Y = true) be π. Let's try to find π given a training set $D = (y_1, y_2, \ldots, y_N)$ with N independent samples of Y. Furthermore, suppose p of the N are positive and n of the N are negative. Write down an expression for the likelihood of D (the probability of seeing this particular sequence of examples, given a fixed value of π) in terms of π, p, and n.
2) By differentiating the loglikelihood L, find the value of π that maximizes the likelihood.
3) Now suppose we add in k Boolean random variables $X_1, X_2, \ldots, X_k$ (the 'attributes') that describes each sample, and suppose we assume that the attributes are conditionally independent of each other given the goal Y. Draw the Bayes net corresponding to this assumption.
4) Write down the likelihood for the data including the attributes, using the following additional assumption:
   - α_i is $P(X_i = true|Y = true)$
   - β_i is $P(X_i = true|Y = false)$
   - p_i^+ is the count of samples for which X_i = true, Y = true.
   - n_i^+ is the count of samples for which X_i = false, Y = true.
   - p_i^- is the count of samples for which X_i = true, Y = false.
   - n_i^- is the count of samples for which X_i = false, Y = false.

   [Hint: consider first the probability of seeing a single example with specified values for $X_1, X_2, \ldots, X_k, Y$]
5) By differentiating the loglikelihood L, find the values α_i and β_i (in terms of various counts) that maximize the likelihood and say in words what these values represent.
6) Let k = 2, and consider a dataset with four examples as follows:

X_1	X_2	Y
0	0	0
0	1	1
1	0	1
1	1	0

   Compute the maximum likelihood estimates of $\pi, \alpha_1, \alpha_2, \beta_1, \beta_2$.
7) Given these estimates of $\pi, \alpha_1, \alpha_2, \beta_1, \beta_2$, what are the posterior probabilities $P(Y = true|X_1, X_2)$ for each example.

III. QUESTION (30 POINTS)

Consider the following training examples, where four Boolean-valued examples (X_1 through X_4) are used to predict a Boolean-valued output (Y1):

X_1	X_2	X_3	X_4	Y
0	1	1	0	1
1	0	1	0	0
1	0	0	1	1
0	1	1	1	0

1) Draw a perceptron for this task, initializing all the weights and biases to -0.1. Show how the weights and biases in this perceptron would be changed after processing the first training example above. Use a learning rate of 0.5. Be sure to show the learning rule you used; dont just show the new weights.
2) Assume now that the weights in your perceptron are all frozen at the value 1, and the only free (ie, adjustable) parameter is the bias (threshold) of the output unit. Draw and briefly explain the weight space for this case (ie, the x-axis is the setting of the bias).
3) Briefly explain why a perceptron cannot learn the "exclusive OR" function?
4) Hand wire (i.e., manually configure) a perceptron so that it computes the Boolean AND of its two inputs.
5) Derive the weight-change rule for a perceptron where the following is the error function and the activation function of the output unit is the identity function (i.e., the output equals the incoming weighted sum and no threshold is involved). Error = output - 5.

IV. QUESTION (24 POINTS)

Consider learning a decision tree that you could use to judge whether or not you will like a given restaurant. Assume you have chosen to use the following three features to describe restaurants, with the possible values shown.

Price $\in$ {Low, Med, High} Type $\in$ {Hamburgers, Pizza, Fish, Vegetarian}

Assume Quinlans ID3 algorithm is given the following set of classified training examples. Calculate the decision tree that ID3 would produce. Show all your work. (You may use the abbreviations that are used to describe the examples.)

P	T	Like
L	H	+
L	V	+
M	F	-
M	V	+
H	P	-

$\log 2 = 1, \log 3 = 1.58, \log 4 = 2, \log 5 = 2.32.$