COMS 4771 HW4

Due: Tue Jun 30, 2015

A printed copy of the homework is due at 5:30pm in class. You must show your work to receive full credit.

1 **[MLE practice]** Consider the data generation process for observation pair (a, b) as follows:

- a is the outcome of an independent six-faced (possibly loaded) dice-roll. That is, chance of rolling face '1' is p_1, rolling face '2' is p_2, etc., with a total of six distinct possibilities.
- Given the outcome a, b is drawn independently from a density distributed as $q_a e^{-q_a x}$ (where $q_a > 0$).

(i) List all the parameters of this process. We shall denote the collection of all the parameters as the variable θ (the parameter vector).

(ii) Suppose we run this process n times independently, and get the sequence:

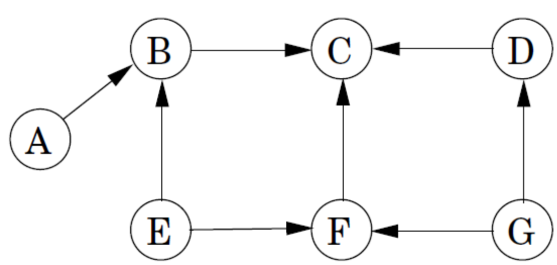
$$(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n).$$

What is the likelihood that this sequence was generated by a specific setting of the parameter vector θ?

(iii) What is the most likely setting of the parameter vector θ given the observation sequence? that is, find the Maximum Likelihood Estimate of θ given the observations.

2 **[Directed graphical models]**

(a) Consider the following directed graphical model:



(i) Which variables are independent of A?

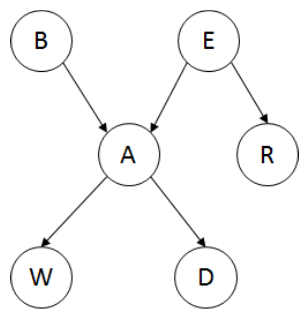
(ii) Which variables are independent of D?

(iii) Which variables are independent of D given F?

(iv) Which variables are independent of D given C?

(v) Define random variables $X = (A, B, E)$, $Y = (C, F)$, $Z = (D, G)$. Draw a directed model which correctly represents the dependencies between these variables. (It should have as few edges as possible and three nodes)?

(b) Consider the following network over six binary variables.



The semantics of this network are as follows. The alarm (A) in your house can be triggered by two possible events: a burglary (B), or an earthquake (E). If there is a strong enough earthquake, there may be a news report (R). If the alarm is ringing, your neighbor Watson calls (W) or your daughter calls (D) you (if they happen to hear the alarm), they may call you even if the alarm is not ringing just to say 'hi'.

(i) Give a simple expression for the joint distribution $P\big[(A, B, D, E, R, W)\big]$.

Given the probability functions: $P(E) = 0.01$, $P(b) = 0.0001$, and

E	$P(R=1\|E)$
0	0.0
1	0.4

A	$P(D=1\|A)$
0	0.0
1	0.7

A	$P(W=1\|A)$
0	0.1
1	1.0

B	E	$P(A=1\|B,E)$
0	0	0.01
0	1	0.2
1	0	0.95
1	1	0.96

(ii) What is the probability that Watson will call?

(iii) What is the probability of a burglary, given that Watson called but the daughter didn't?

(iv) What is the probability of an earthquake, given that there was no news report, but both Watson and the daughter called?

(v) What is the most likely explanation of the following scenario: Watson doesn't call, daughter calls, and there is no news report?

(c) The notation "$A \perp B \mid C$" means "A and B are independent given C". Show that:

$$X \perp Y \mid W, Z \text{ and } X \perp W \mid Z \implies X \perp W, Y \mid Z.$$

3 **[From distances to embeddings]** Your friend from overseas is visiting you and asks you the geographical locations of popular US cities on a map. Not having access to a US map, you realize that you cannot provide your friend accurate information. You recall that you have access to the relative distances between nine popular US cities, given by the following distance matrix D:

Distances (D)	BOS	NYC	DC	MIA	CHI	SEA	SF	LA	DEN
BOS	0	206	429	1504	963	2976	3095	2979	1949
NYC	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIA	1504	1308	1075	0	1329	3273	3053	2687	2037
CHI	963	802	671	1329	0	2013	2142	2054	996
SEA	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DEN	1949	1771	1616	2037	996	1307	1235	1059	0

Being a machine learning student, you believe that it may be possible to infer the locations of these cities from the distance data. To find an embedding of these nine cities on a two dimensional map, you decide to solve it as an optimization problem as follows.

You associate a two-dimensional variable x_i as the unknown latitude and the longitude value for each of the nine cities (that is, x_1 is the lat/lon value for BOS, x_2 is the lat/lon value for NYC, etc.). You write down the an (unconstrained) optimization problem

$$\text{minimize}_{x_1,\ldots,x_9} \quad \sum_{i,j} \big(\|x_i - x_j\| - D_{ij}\big)^2,$$

where $\sum_{i,j}(\|x_i - x_j\| - D_{ij})^2$ denotes the embedding discrepancy function.

(i) What is the derivative of the discrepancy function with respect to a location x_i?

(ii) Write a program in your preferred language to find an optimal setting of locations $x_1, \ldots, x_9$.

(iii) Plot the result of the optimization showing the estimated locations of the nine cities. (here is a sample code to plot the city locations in Matlab)

```
>> cities={'BOS','NYC','DC','MIA','CHI','SEA','SF','LA','DEN'};
>> locs = [x1;x2;x3;x4;x5;x6;x7;x8;x9];
>> figure; text(locs(:,1), locs(:,2), cities);
```

What can you say about your result of the estimated locations compared to the actual geographical locations of these cities?