EECS545 - Homework 5

March 15th, 2018

Instructions

- This homework is Due Tuesday, March the 27th at 5pm. No late submissions will be accepted.
- As always, submit a write-up and your code for this homework.
- You are expected to use Python for the programming questions in this homework. Use of Python 3 is recommended.
- **Submit your write-up to Gradescope under the assignment titled** `Homework 5`. Make sure to tag the pages with the corresponding problems numbers.
- Submit all your python code to Canvas in a compressed zip file named `uniqname_hw5_code.zip`. Your zip file should contain `prob4.py`.

1. Let $\{A, B, C, D\}$ be binary random variables. Consider the following independence assumptions.

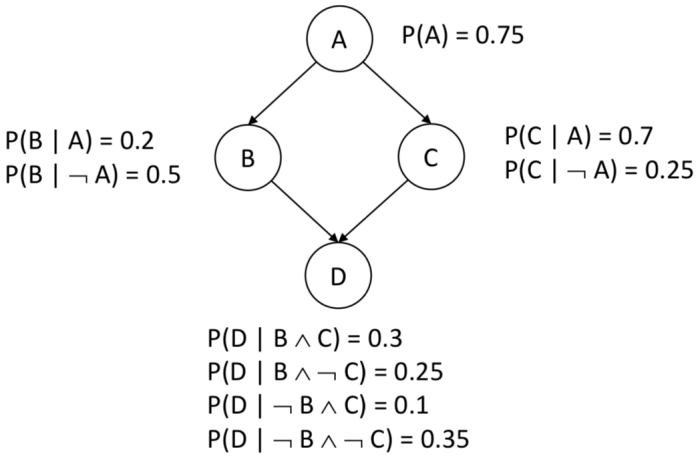
   - $A \perp B$
   - $A \not\perp D|B$
   - $A \perp D|C$
   - $A \not\perp C$
   - $B \not\perp C$
   - $A \not\perp B|D$
   - $B \perp D|A, C$

   where $\perp$ represents conditionally independent and $\not\perp$ means not conditionally independent.

   (a) Draw a Bayes net over the variables $\{A, B, C, D\}$ where the above independence assumptions hold (You are expected to draw a single Bayes net).

   (b) How many parameters are needed to define the Conditional Probability Distributions (CPD's) for this Bayes Net?

2. Consider the following Bayesian network where A, B, C, D are Boolean random variables.



(a) Write down all the conditional independencies in the above Bayes Net.

(b) Construct the probability table for $P(B, C, D|\neg A)$.

3. Consider the following probability table over binary random variables A, B, C. Draw all possible Bayes nets which could represent the distribution over these variables.

A	B	C	P(A,B,C)
0	0	0	0.03
0	0	1	0.045
0	1	0	0.12
0	1	1	0.105
1	0	0	0.035
1	0	1	0.14
1	1	0	0.315
1	1	1	0.21

4. **Gaussian Mixture Models.** In this problem you will fit GMMs to a synthetic dataset. The data is provided in a $N \times 2$ numpy array (N data-points) named `gmm_data.npy`.

(a) Implement the EM algorithm to fit a GMM to the given data. Initialize the model parameters appropriately. For each of the following values of $K \in \{2, 3, 5, 10\}$ (the number of Gaussians) provide a plot consisting of the following:

- A scatter plot of the data
- The learned Gaussian components in the mixture

Note: The learned parameters can be sensitive to the parameter initialization. So you may have to try different initialization schemes to obtain acceptable results.

(b) Now fit a GMM with $K = 3$ using the parameter initialization provided in the starter code `prob4.py`. The initialization parameters are provided in three lists `pi,means,covs` whose elements correspond to π_k, μ_k, Σ_k respectively (assuming the standard notation). Perform 50 steps of EM. Report the following:

- Provide a sequence of plots similar to the one you provided above after 1, 5, 10, 20, 50 steps of the algorithm (Optionally, color code the points by the responsibility scores).

- Report the model parameters learned by EM after 50 steps.

(c) (Ungraded) Explore the effect of different parameter initializations on the behavior of the learning algorithm.