Homework-4

COMP7745/8745

FALL 2018

Instructor: Deepak Venugopal

Due Date: April 25 (before class)

1)When the clusters are well-defined, which algorithm would you prefer among k-Means and Gaussian mixture Models, and why? (10 points)

2)Suppose someone tells us in advance that there are k clusters in a dataset, can we guarantee that K-Means will find those clusters, if we initialize it with k cluster-centers? Briefly Explain (10 points)

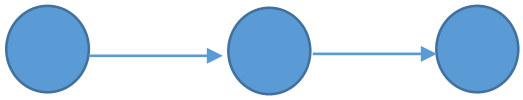
3) Suggest some ideas as to how could we use clustering for detecting anomalies in a dataset? (10 points)

4) You run the EM algorithm for Gaussian Mixture Models, and obtain the following weight matrix upon convergence. Briefly describe in your own words, what this matrix suggests in terms of the clustering. (10 points)

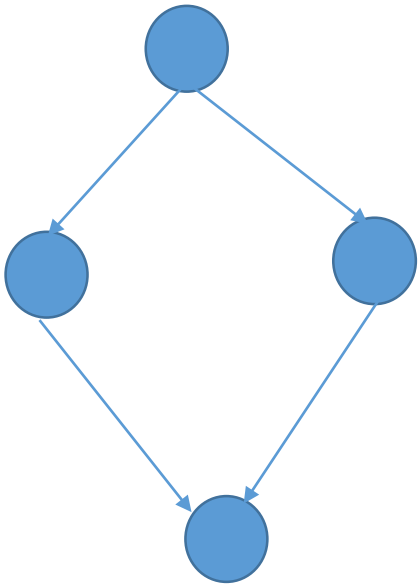
$$\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.7 & 0.1 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

5) How many parameters would we need to learn for each of the following Bayesian networks (assuming binary random variables corresponding to each node) (10 points)

a)

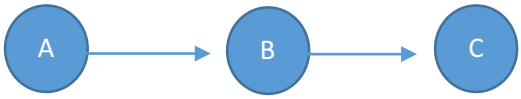


b)

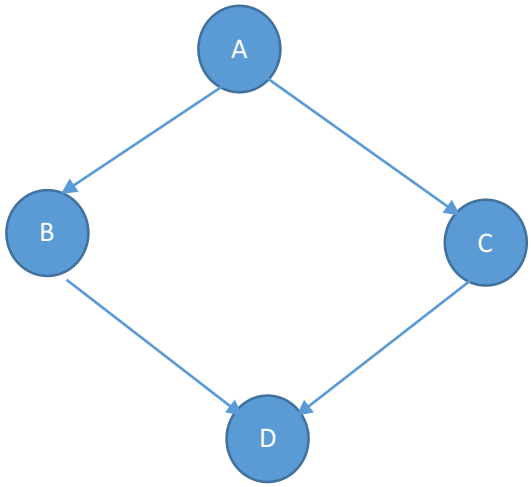


6) In the following Bayesian networks, state all the conditional independencies that are specified by the network structure (10 points)

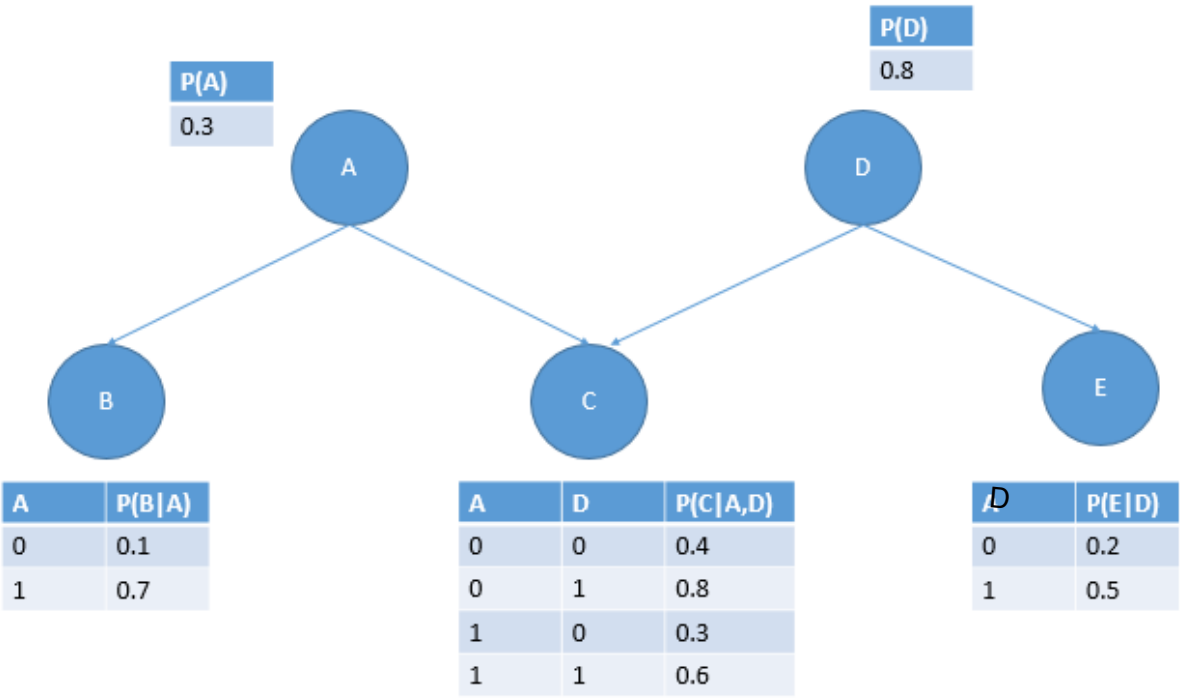
a)



b)



7)Consider the following Bayesian network (15 points)



A	P(B|A)
0	0.1
1	0.7

A	D	P(C|A,D)
0	0	0.4
0	1	0.8
1	0	0.3
1	1	0.6

D	P(E|D)
0	0.2
1	0.5

Answer the following queries based on the above network

a) P(A=0,B=1,C=0,D=1,E=0)
b) P(C=1|A=0,D=1)
c) P(E=1|A=0,B=0,C=1,D=1)