Assignment 7

Problem 1

Define

$$J_1(\boldsymbol{w}) = |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}|^2 + \lambda_2|\boldsymbol{w}|^2 + \lambda_1|\boldsymbol{w}|_1$$

and

$$J_2(\boldsymbol{w}) = |\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{w}|^2 + c\lambda_1|\boldsymbol{w}|_1$$

where $c = (1+\lambda_2)^{-\frac{1}{2}}$ and

$$\tilde{\boldsymbol{X}} = c\begin{pmatrix}\boldsymbol{X}\\ \sqrt{\lambda_2}\boldsymbol{I}_d\end{pmatrix},\ \tilde{\boldsymbol{y}} = \begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{0}_{d\times 1}\end{pmatrix}$$

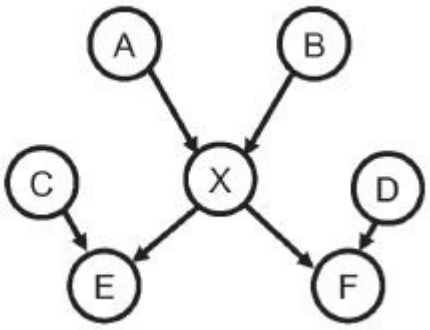
Show

$$\arg\min J_1(\boldsymbol{w}) = c(\arg\min J_2(\boldsymbol{w}))$$

i.e. $J_1(c\boldsymbol{w}) = J_2(\boldsymbol{w})$ and hence that one can solve an elastic net problem using a lasso solver on modified data.

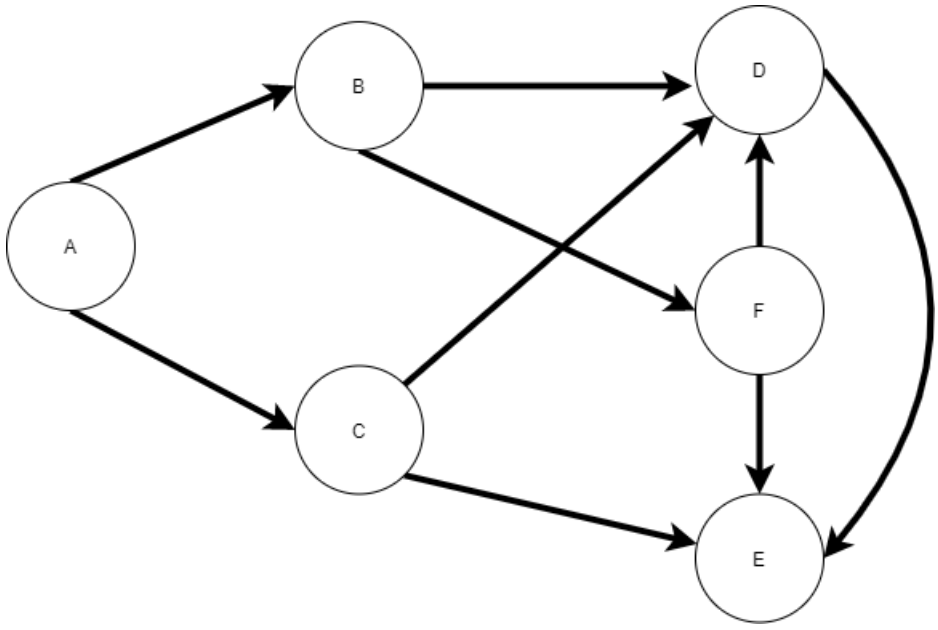
Problem 2

Consider the DAG G in the following figure. Assume it is a minimal I-map for $p(A, B, C, D, E, F, X)$. Now consider marginalizing out X. That is, X is unobserved and we construct $p(A, B, C, D, E, F) = \sum_X p(A, B, C, D, E, F, X)$. Construct a new DAG G' which is a minimal I-map for $p(A, B, C, D, E, F)$. Specify (and justify) which extra edges need to be added. Discuss which pairs of variables are conditionally independent in your new graph.



Assignment 7

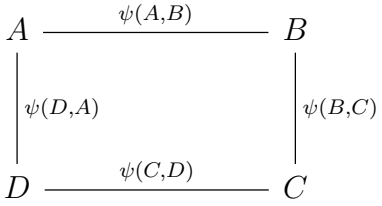
Problem 3



1. Convert the BN above to MRF.
2. Fill in the blanks to make a true statement: _____ and _____ are d-separated given _____ in the BN, but not d-separated in the MRF.

Problem 4

Consider a MRF of 4 random binary variables A, B, C, D, corresponding to 4 students in a CMPT 727 study group. Each variable represents whether the student has a correct understanding of the EM algorithm.



The potential function tables are defined as follows:

(var1, var2)	$\psi(A, B)$	$\psi(B, C)$	$\psi(C, D)$	$\psi(D, A)$
(0,0)	100	10	1	10
(0,1)	1	1	10	1
(1,0)	1	1	10	1
(1,1)	100	50	1	20

This table is formed based on observation that A and B are good friends, so they are prone to the same understanding (either correct or incorrect); two other pairs (B,C) and (D,A) also tend to have same understanding but they are more likely be correct. C and D usually argue, and they are more likely to hold different opinions.

1. Calculate the partition function Z.
2. What is the probability that all students in this group have a correct understanding of the EM algorithm?

Problem 5

Please write one thing from this course you found confusing, a topic you would like to hear more about, or something you found particularly interesting.