COMP 7745/8745: Final Exam

May 2, 2017

Allocated Time: 2 hour and 00 minutes

NAME: ----------

- **Please write your answers in the sheets provided**
- **State any assumptions if you need to make them.**
- **Best of luck!**

Maximum Points: 100

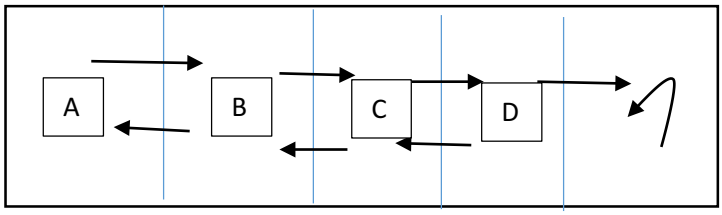
Number of pages: 5

1) Answer the following true/false questions with a brief justification (39 points) (no points will be given for random guessing without justification or for wrong justification)

    a. The K-Means algorithm will always find the optimal number of clusters
    b. Given that the size of the hypothesis space H is n, vc-dimension(H) ≤ n
    c. The main advantage of SVMs is that they can work with unlabeled data
    d. Kernels typically reduce the dimensionality of the feature-space to improve accuracy
    e. The VC dimension of a classifier is independent of the dataset used by the classifier
    f. A rough estimate of the leave-one-out cross validation error of a SVM classifier with k/4 support vectors and k examples is equal to 0.25
    g. Q-learning converges faster if immediate rewards are more important than deferred rewards

h. If our data has clearly demarcated clusters, it is generally better to use KMeans instead of Gaussian Mixture Models for clustering
i. Q-Learning is useful in learning a policy that is optimal for a specific starting state
j. The computational complexity when using a polynomial kernel of degree 100 is significantly more than the complexity when using a polynomial kernel of degree 2.
k. Learning the parameters of Bayesian network where the data is fully-observed is typically harder than learning the parameters when some data is missing.
l. The VC-Dimension of a linear kernel SVM is smaller than the VC-dimension of an exponential kernel SVM.
m. The number of support vectors in an SVM depends upon both training data and the choice of kernel.

2) Short answers (9 points)
   a. Consider the following distribution: P(A,B,C,D) = P(D)*P(B)*P(A)*P(C|A,B,D). How many parameters will the corresponding Bayesian network have?
   b. Consider the following distribution: P(A,B,C.D) = P(D)*P(C)*P(B)*P(A). How many edges will the corresponding Bayesian network have?
   c. Consider the following board game with rewards indicated on the arrows. Let the discount factor be 0.5.



Given infinite iterations, what will be the following values converge to using the Q-Learning algorithm

Q(D,Right), Q(C,Right), Q(B,Right), Q(A,Right)

3) Support Vector Machines (20 points)

   a. For the following training dataset using a linear kernel, formulate the dual problem that needs to be solved by the SVM.

Feature-1	Feature-2	Label
-1	-1	-1
1	1	1

   b. You formulate the SVM dual problem using a linear kernel and feed it to a blackbox optimizer. The optimizer returned an answer where all the Lagrangian multipliers have a

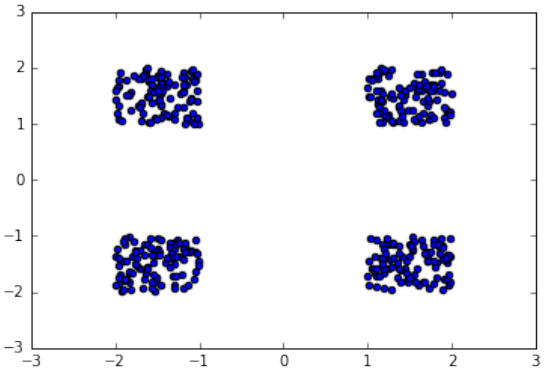
value equal to infinity. (i) What can you gather about the dataset and what is the next logical step that you would take for this dataset? (ii) If you could choose your own dataset, which of the following two datasets will you choose for an SVM based classifier and why?

Dataset A: All Lagrangian multipliers computed on this dataset have values greater than 0

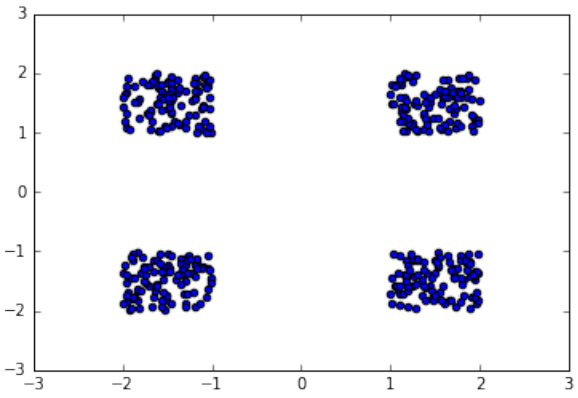
Dataset A: Only half the Lagrangian multipliers computed on this dataset have values greater than 0

4) Computational Learning Theory (12 points)
   a) Suppose our input is a set of n boolean variables and I tell you that the hypothesis space H is the set of all possible conjunctions and all possible disjunctions. E.g. formulas of the form, $X \vee Y \vee \neg Z \vee \ldots$. are conjunctions whereas formulas of the form $X \wedge Y \wedge \neg Z \wedge \ldots$. are disjunctions. What is the minimum number of samples required to learn the correct concept from H such that the generalization error is guaranteed to be less than 0.1 with confidence 95% (A loose sample complexity bound is acceptable as an answer and you don't have to use a calculator to simplify your final result to compute the numerical value)

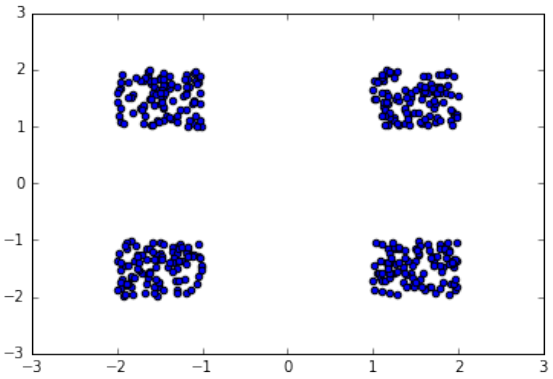
5) We wish to find clusters in the following datasets. In each case, approximately to what centroids will K-means converge to (you don't need to do any elaborate calculations here, you can simply show the centroids in the figure itself). (12 points)
   a. For the following case, assume that we choose K=4 and the initial cluster centers are at the 4 corners. (-3,3); (3,3); (3,-3); (-3,-3)



   b. For the following case, assume that we choose K=2 and the initial cluster centers are at (0,-3); (0,3)



c. For the following case, assume that we choose K=2 and the initial cluster centers are at (0,0); (0,0).



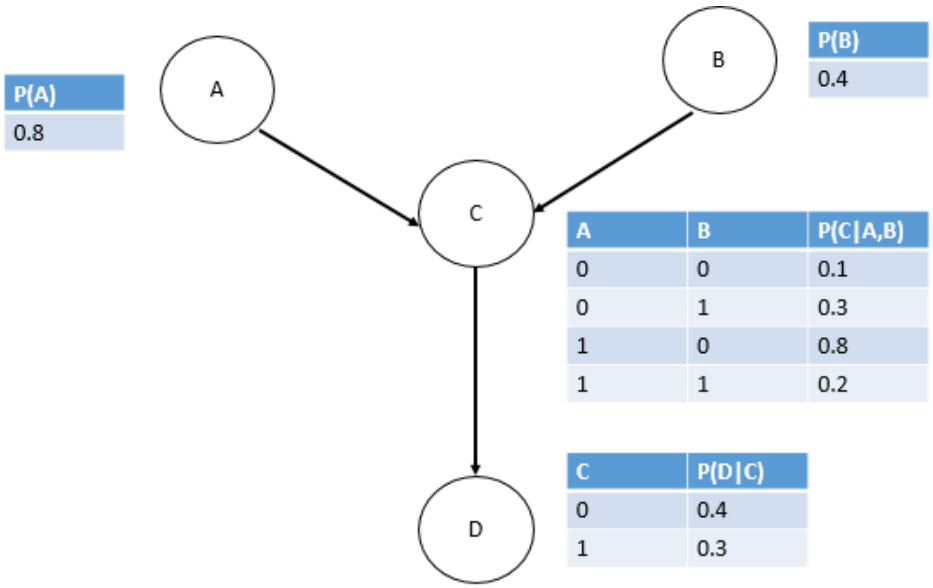
6. Answer the below questions for the given Bayesian network (8 points)

a. P(A=1,B=1,C=1,D=0) = ?

b. P(D=0|A=1,B=1,C=0) = ?

c. Do you think P(A=0) = P(A=0|D=1)?

d. If we want to add a new node E which is conditionally independent of D given C, how can we add this



P(A)
0.8

P(B)
0.4

A	B	P(C|A,B)
0	0	0.1
0	1	0.3
1	0	0.8
1	1	0.2

C	P(D|C)
0	0.4
1	0.3