

COMP 7745/8745: Final Exam

April 27, 2018

Allocated Time: 2 hour and 00 minutes

NAME:

1) Support Vector Machines (12 points)

Consider the following training dataset.

Feature-1	Feature-2	Label
-1	0	-1
1	0	1

- Using a polynomial kernel of degree 3, formulate the dual problem that needs to be solved by the SVM. For a polynomial kernel of degree 3, the kernel function is given by $K(x_i, x_j) = (1 + x_i^T x_j)^3$ where x_i, x_j are two instances in the training data.
- Using Lagrangian multipliers $\alpha_1 = 1, \alpha_2 = 2$, show the computation that needs to be carried out to classify the new instance $(-1, -1)$

2) Suppose Q-learning converges to the values given in the below grid. What is the optimal policy? (10 points)

A	← 10	C
	→ 20	
B	← 20	30 ↓
	→ 40	↑ 50
		D

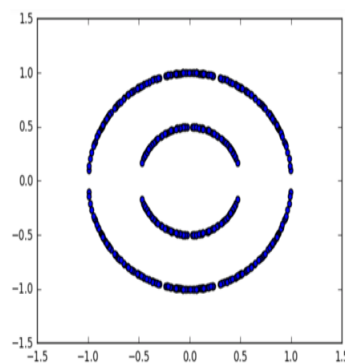
3) Suppose we have the following 1-D dataset, what clusters will K-Means find in each of the following cases? (break ties randomly) (10 points)

Feature
1
2

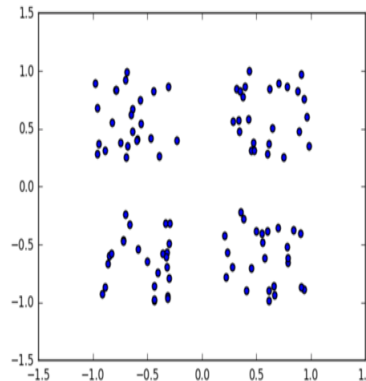
3
4
5
6
7
8
9
10

- a. 2 Cluster centers initialized to 1 and 10 respectively
- b. 2 Cluster centers initialized to 5 and 6 respectively

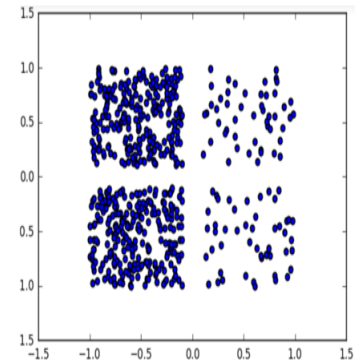
4) Which of the following cases would the K-Means algorithm probably work best on. Briefly explain why? (10 points)



(a)

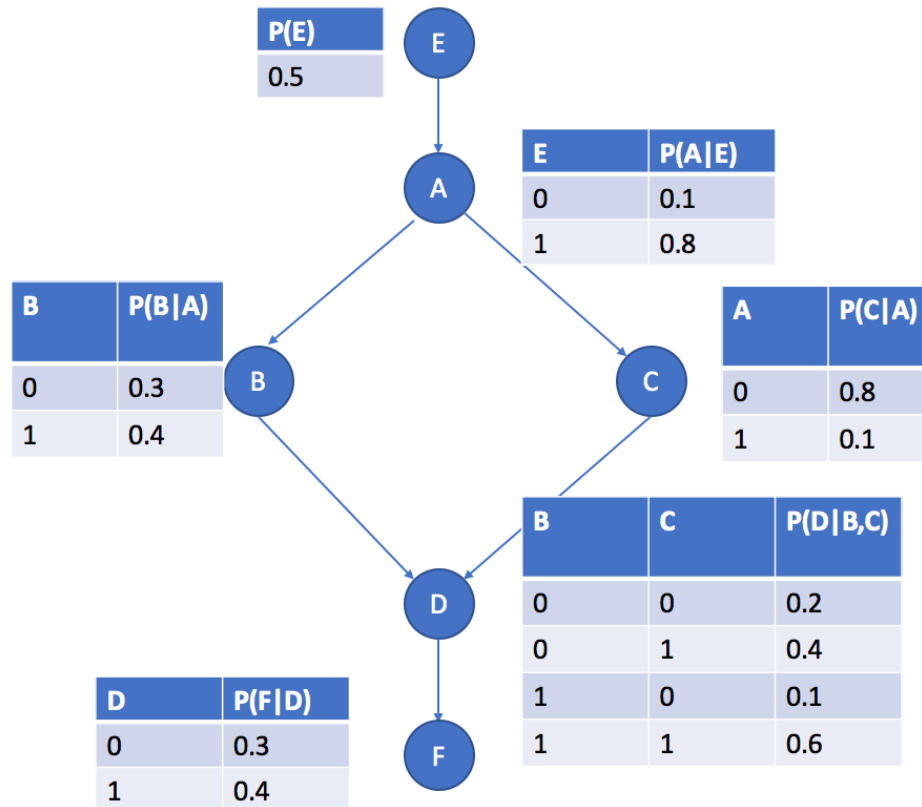


(b)



(c)

- 5) Suppose you train a linear SVM classifier on a dataset, and observe that the accuracy of the classifier is quite low (say around 0.6) on the training dataset. Your boss wants you to improve its accuracy and has heard of kernels and boosting as potential solutions. In your own words, explain to your boss, the pros and cons of using each of these solutions. (10 points)
- 6) Consider the following Bayesian network and answer the given questions (15 points)



- $P(A=0, B=0, C=0, D=1, E=1, F=1)$
- $P(F=0 \mid D=0, A=0, E=0)$
- Is $P(D \mid B, C, F) = P(D \mid B, C)$? Briefly explain.
- What conditional independencies can we infer from the Bayesian network?
- Suppose we get some new information that suggests that we add an edge starting at A and ending at D, how many CPTs need to be modified due to this change? How many parameters would the new Bayesian network have?