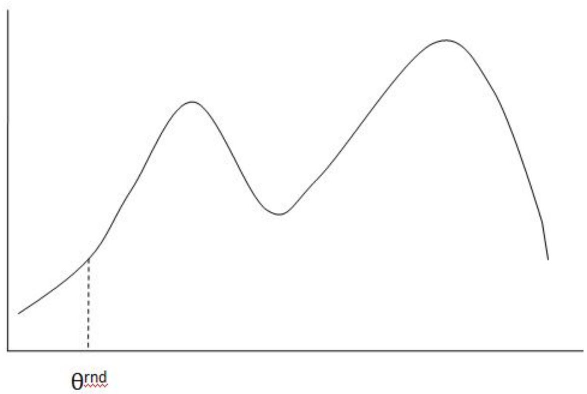
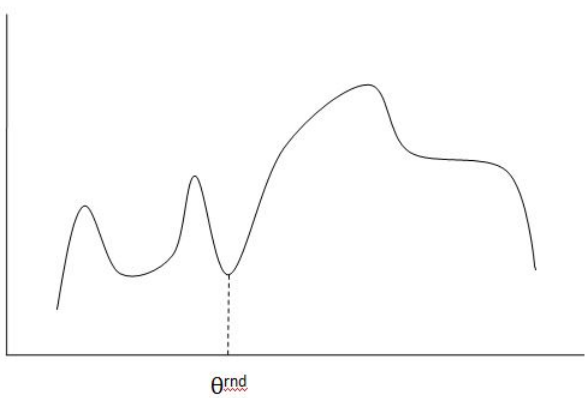
Assignment 11

Introduction to Machine Learning
Prof. B. Ravindran

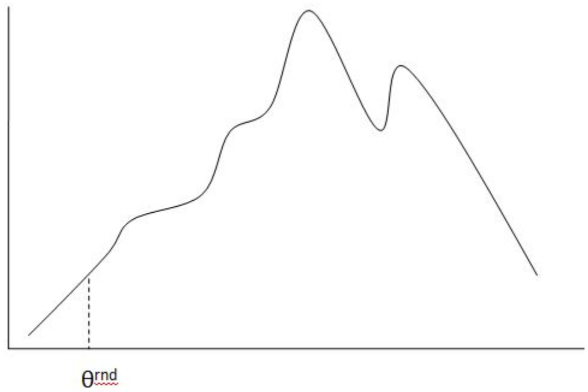
1. In each of the options the diagram contains a log likelihood function for a particular problem as well as an initial value of the parameter used in the execution of the EM algorithm. In which of the given scenarios will the EM algorithm be able to achieve the global maximum?



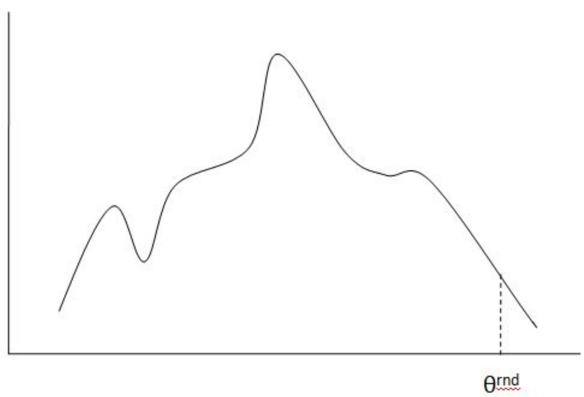
Q1 (a)



Q1 (b)



Q1 (c)



Q1 (d)

2. Suppose we are given n p-dimensional data points and the corresponding class labels (k different classes). We want to build a decision tree classifier to classify the data. However, we find that there are missing values in the data set. Is it possible to use the EM algorithm to fill the missing data given the above information and making no further assumptions?

   (a) Yes

   (b) No

3. Suppose you have a PAC learning algorithm A for a concept class C such that with probability at least 0.5, the algorithm will output an approximately correct hypothesis. Suppose that for deployment purposes, you need an algorithm which can output the approximately correct hypothesis with probability at least 0.998. Is it possible to make use of algorithm A for this purpose? If so, how?

   (a) No, we cannot make use of A

   (b) Yes, repeat A three times and choose the best hypothesis

(c) Yes, repeat A five times and choose the best hypothesis

(d) Yes, repeat A nine times and choose the best hypothesis

4. To say that the VC-dimension of a class is at least k, is it necessary for the class to be able to shatter any configuration of k points?

   (a) Yes

   (b) No

5. What is the VC-dimension of the class of axis-parellel rectangles?

   (a) 3

   (b) 4

   (c) 5

   (d) 6

6. Which of the following statements are true about similarity graph based representations which are used for spectral clustering? (Note that more than one statements may be correct)

   (a) One can give a tighter upper bound than $O(n)$ (where n is the number of data points) on the maximum degree of the vertex corresponding to a point in its kNN based similarity graph representation

   (b) One can give a tighter upper bound than $O(n)$ (where n is the number of data points) on the maximum degree of the vertex corresponding to a point in its epsilon neighborhood based similarity graph representation

   (c) If a is in the k nearest neighbors of b, then b is in the k nearest neighbors of a

   (d) If a is in the epsilon neighborhood of b, then b is in the epsilon neighborhood of a