1. Explain what principal component analysis does?

PCA is a technique used to emphasize variation and bring out strong patterns in dataset.

In a data set with dimension N, the main dimension variation will be a N-dimension line. Then the left N-1 dimensions are orthogonal to the line. And we can represent out data in the N-1 dimension while container its main variations. The removed 1 dimension has least variation on that dimension. So as long as we repeat this process, we can reduce the N dimension data set into M dimension, (M is less then N), then in the less dimension space, we still have the most variation on data for us to process it with mush less effort.

1. Explain how k-means clustering works?

1) Input: K, set of points X1 to Xn

2) Place centroids C1 … Ck at random locations

* For each point X:

Find nearest centroid Cj

Assign the points Xi to cluster j

* For each cluster j = 1 … K:

NEW Cj 🡨 compute the mean of all points in cluster j, Xj

Repeat previous, for each point X, assign point Xi to cluster j.

3) Stop when moving distance from previous Cj to new Cj less then some threshold.

1. Explain what an autoassociative network is?

We build multiple layer perceptron neural network with one hidden layer, and train it to map each training item onto itself. The hidden layer is smaller then the input/out layers. Thus, the hidden layer compresses the data from input. This kind of neural network is called autoassociative network.

1. Quiz Question:
2. What order would the nodes be encountered if the graph was searched

n1, n2, n4, n5, n3, n6, n7 (depth first, left to right)

n1, n3, n2, n7, n6, n5, n4 (breath first, right to left)

1. For each of the above search strategies, define an algorithm that searches the graph using the notion of a **fringe** of nodes to be expanded.

Fringe stack is denoted as ().

For depth first, left to right:

1. (n1) n1-> n2, n3
2. (n2, n3) n2-> n4, n5
3. (n4, n5, n3)
4. (n3) n3-> n6, n7
5. (n6, n7)

For breath first, right to left:

1. (n1) n1-> n3, n2
2. (n3, n2) n3-> n7, n6
3. (n2, n7, n6) n2-> n5, n4
4. (n7, n6, n5, n4)