gradiance

Gradiance Online Accelerated Learning

Robert

- Home Page
- Handouts
- Tutorials
- Homeworks
- Lab Projects
- Reports
- Class Administration
- Question Bank
- Log Out

Help

03: SVD and Clustering

Number of questions:	5
Positive points per question:	3.0
Negative points per question:	1.0

Gradiance quiz on Dimensionality Reduction and Clustering. You can attempt to answer the questions as many times as you like. Questions get randomly regenerated each time. The score of the *last* submission gets saved into our records (that is, once you get a perfect score, don't submit again with a bad one).

1. Note: In this question, all columns will be written in their transposed form, as rows, to make the typography simpler. Matrix M has three rows and three columns, and the columns form an orthonormal basis. One of the columns is [2/7,3/7,6/7], and another is [6/7, 2/7, -3/7]. Let the third column be [x,y,z]. Since the length of the vector [x,y,z] must be 1, there is a constraint that $x^2+y^2+z^2 = 1$. However, there are other constraints, and these other constraints can be used to deduce facts about the ratios among x, y, and z. Compute these ratios, and then identify one of them in the list below.

a) z = 3y b) 2x = 3z c) y = 2x d) y = 3z

2. In certain clustering algorithms, such as CURE, we need to pick a representative set of points in a supposed cluster, and these points should be as far away from each other as possible. That is, begin with the two furthest points, and at each step add the point whose minimum distance to any of the previously selected points is maximum.

Suppose you are given the following points in two-dimensional Euclidean space: x = (0,0); y = (10,10), a = (1,6); b = (3,7); c = (4,3); d = (7,7), e = (8,2); f = (9,5). Obviously, x and y are furthest apart, so start with these. You must add five more points, which we shall refer to as the first, second,..., fifth points in what follows. The distance measure is the normal Euclidean L_2-norm. Which of the following is true about the order in which the five points are added?

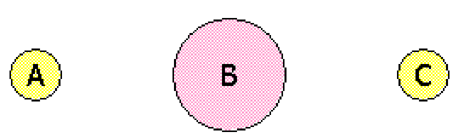
a) b is added first

b) b is added second

c) e is added third

d) a is added fourth

3. Suppose that the true data consists of three clusters, as suggested by the diagram below:



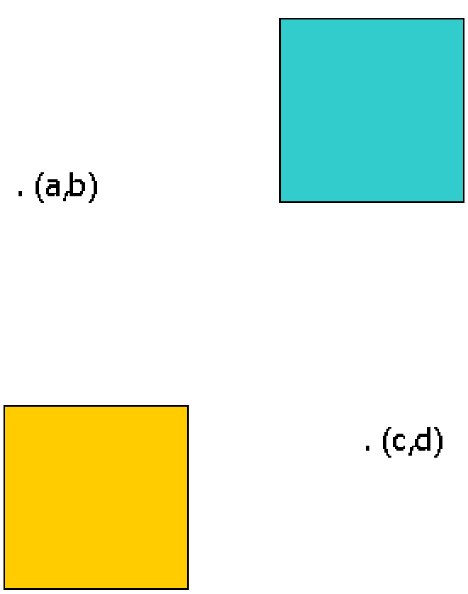
There is a large cluster B centered around the origin (0,0), with 8000 points uniformly distributed in a circle of radius 2. There are two small clusters, A uniformly distributed in a circle of radius 2. There are two small clusters, A and C, each with 1000 points uniformly distributed in a circle of radius 1. The center of A is at (-10,0) and the center of C is at (10,0).

Suppose we choose three initial centroids x, y, and z, and cluster the points according to which of x, y, or z they are closest. The result will be three *apparent* clusters, which may or may not coincide with the *true* clusters A, B, and C. Say that one of the true clusters is *correct* if there is an apparent cluster that consists of all and only the points in that true cluster. Assuming initial centroids x, y, and z are chosen independently and at random, what is the probability that A is correct? What is the probability that C is correct? What is the probability that both are correct?

After computing these probabilities, identify the true statement from the list below. Note: probabilities are rounded to the nearest percent.

- a) The probability that A is correct is 19%
- b) The probability that both A and C are correct is 8%
- c) The probability that C is correct is 5%
- d) The probability that A is correct is 24%

4. When performing a k-means clustering, success depends very much on the initially chosen points. Suppose that we choose two centroids (a,b) = (5,10) and (c,d) = (20,5), and the data truly belongs to two rectangular clusters, as suggested by the following diagram:



Under what circumstances will the initial clustering be successful? That is, under what conditions will all the yellow points be assigned to the centroid (5,10), while all of the blue points are assigned to cluster (20,5))? Identify in the list below, a pair of rectangles (described by their upper left corner, UL, and their lower-right corner LR) that are successfully clustered.

- a) Yellow: UL=(7,8) and LR=(12,5); Blue: UL=(15,14) and LR=(20,10)
- b) Yellow: UL=(6,7) and LR=(11,4); Blue: UL=(11,5) and LR=(17,2)
- c) Yellow: UL=(6,7) and LR=(11,4); Blue: UL=(14,10) and LR=(23,6)
- d) Yellow: UL=(3,3) and LR=(10,1); Blue: UL=(13,10) and LR=(16,4)

5. We want to do an approximate UV-decomposition of the matrix M =

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

We shall use only a single column for U and a single row for V, so the goal is to make the product UV as close as possible to M. Initially, we shall set V to [5,5,5] and make the entries of U unknown. Then in the first step, we choose the values of x, y, and z that minimize the root-mean-square error (RMSE) between the product

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} 5 & 5 & 5 \end{bmatrix}$$

and the matrix M.

Find the values of x, y, and z that minimize the RMSE and identify one of those values below.

- a) z = 3/5.
- b) y = 1.
- c) y = 2.
- d) y = 5.