Diep Vu

Homework 1

Part I

Basic Concepts in Linear Algebra and Calculus

1. We have two vectors, x_1 and x_2

$x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 10 \\ 18 \end{bmatrix}$

What is the distance between x_1 and x_2 ?

(1) if the distance measure is based on L2 norm (a.k.a Euclidean norm)

(2) if the distance measure is based on L1 norm

(3) if the distance measure is based on L∞ norm (a.k.a infinity norm)

Assuming there are two feature components $x = \begin{bmatrix} income \\ spend \end{bmatrix}$ in an application, does the L∞ norm-based distance measure make sense for the application of customer segmentation?

2. We define a scalar valued function of a vector variable

$$f(x) = x^T A x$$

Here, x is a column vector, x^T is the transpose of x, and A is a symmetric matrix

To simplify this question, let's assume x has only two elements $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, and $A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$

The derivative of f with respect to x is a vector defined by $\frac{df}{dx} = \begin{bmatrix} \frac{df}{d\alpha} \\ \frac{df}{d\beta} \end{bmatrix}$

Show that $\frac{df}{dx} = 2Ax$

Hint: calculate $f(x)$, $2Ax$, $\frac{df}{d\alpha}$ and $\frac{df}{d\beta}$

K-means clustering

3. Briefly describe the two key steps in one iteration of the k-means algorithm.

4. What is the distance measure used in k-means (implemented in sk-learn)?

5. The k-means algorithm can converge in a finite number of iterations. Why?

6. The clustering result of k-means could be random. Why?

7. The minimum value of the objective/loss function is zero for any dataset. What is the clustering result when the objective function is zero?

Note: for questions 3,4,5,6,7, you only need to write a few words (bullet points) for each one.

You may write the answers on a piece of paper, take a photo using your cell phone, and upload the picture to Blackboard. **Make sure that your handwriting is human-readable**.

You may use MS-word to write the answers, convert the file to PDF, and upload it to Blackboard.

Part 2: Programming

Complete the tasks in the files:

H1P2T1_kmeans.ipynb

If you want to get some bonus points, try this task:

H1P2T2_kmeans_compression.ipynb

Grading: the number of points

	Undergraduate Student	Graduate Student
Basic Concepts in Linear Algebra and Calculus	10	10
K-means clustering	10	10
H1P2T1	30	30
H1P2T2	10 (bonus)	10 (bonus)
Total number of points	50 + 10	50+10

Upload your files (*_your_name.ipynb) to blackboard

Do NOT covert the ipynb files to pdf.

Basic Concepts in Linear Algebra and Calculus

1. $x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 10 \\ 18 \end{bmatrix}$

Find the distance between x_1 and x_2 based on

a. L2 norm (Euclidean norm)

$$d = \sqrt{(1-10)^2 + (2-18)^2}$$

$$= \sqrt{9^2 + 16^2} = \sqrt{337} \simeq 18.36$$

b. L1 norm

$$d = |1-10| + |2-18|$$

$$= 9 + 16 = 25$$

c. $L\infty$ norm

$$d = \max(|1-10|, |2-18|)$$

$$= \max(9, 16) = 16$$

d. $L\infty$ norm-based distance measure only considers the maximum absolute value between 2 vectors, not the relationship between 2 features

↳ It doesn't make sense to use L-∞ norm-based distance

It would make more sense to use the L2 norm

2. $f(x) = x^T A x$

$x^T A x$: x → column vector; A → symmetric matrix

$x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ $\quad A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$

$$\frac{df}{dx} = \begin{bmatrix} \frac{df}{d\alpha} \\ \frac{df}{d\beta} \end{bmatrix}$$

Show that $\frac{df}{dx} = 2Ax$

$$f(x) = x^T A x = \begin{bmatrix} \alpha & \beta \end{bmatrix} \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$= \begin{bmatrix} \alpha a + \beta c & \alpha c + \beta b \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$= \alpha^2 a + \alpha \beta c + \alpha c \beta + \beta^2 b = a\alpha^2 + 2c\alpha\beta + b\beta^2$$

$$2Ax = 2\begin{bmatrix} a & c \\ c & b \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = 2\begin{bmatrix} a\alpha + c\beta \\ c\alpha + b\beta \end{bmatrix}$$

$$\left.\begin{aligned} \frac{df}{d\alpha} &= 2a\alpha + 2\beta c \\ \frac{df}{d\beta} &= 2c\alpha + 2b\beta \end{aligned}\right\} \rightarrow \frac{df}{dx} = \begin{bmatrix} 2a\alpha + 2\beta c \\ 2c\alpha + 2b\beta \end{bmatrix} = 2\begin{bmatrix} a\alpha + c\beta \\ c\alpha + b\beta \end{bmatrix} = 2Ax$$

K-means Clustering

3. Two key steps in one iteration of the k-means algorithm
   - Update centers: for each cluster, move the center vector c to the average location of the data points in the cluster
   - Update labels: for each data point, find the nearest cluster center and attach a cluster label to the data point.
4. K-means uses ℓ_2 norm (Euclidean norm)
5. K-means can coverge in a finite number of iterations since it is based on an optimization problem whose objective function is to minimize the loss. After a number of iterations, the loss curve becomes flat. The objective function would reach a minimum value, which is the global optimum
6. The clustering result of k-means would be random since clustering result is determined by data distribution and initialization, and the initialization of the centers is random.
7. When the object function is zero, the clustering result is very good, meaning that the cluster assignments are good and the clusters are well separated.