

CSC546:Machine Learning and its Applications

Homework 1

Aaron Jesus Valdes

February 8, 2021

Part I

1

a

$$x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad x_2 \begin{bmatrix} 10 \\ 18 \end{bmatrix} \quad (1)$$

$$||x - y||_2 = \sqrt{(x[1] - y[1])^2 + (x[2] - y[2])^2} \quad (2)$$

$$||x - y||_2 = \sqrt{(1 - 10)^2 + (2 - 18)^2} \quad (3)$$

$$||x - y||_2 = \sqrt{337} \approx 18.3575 \quad (4)$$

(5)

b

$$x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad x_2 \begin{bmatrix} 10 \\ 18 \end{bmatrix} \quad (6)$$

$$L_1 = (x[1] - y[1]) + (x[2] - y[2]) \quad (7)$$

$$L_1 = |1 - 10| + |2 - 18| \quad (8)$$

$$L_1 = 25 \quad (9)$$

(10)

c

$$x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad x_2 \begin{bmatrix} 10 \\ 18 \end{bmatrix} \quad (11)$$

$$L_1 = \max(\max(|x[1] - y[1]|) \max(|x[2] - y[2]|)) \quad (12)$$

$$L_1 = \max(\max(|1 - 10|) \max(|2 - 18|)) \quad (13)$$

$$L_1 = \max(\max(|9|) \max(|16|)) \quad (14)$$

$$L_1 = \max(16) \quad (15)$$

$$L_1 = 16 \quad (16)$$

(17)

d

The L_{inf} does not make sense for these two features since this can lead to the loss of essential features for customer segregation given the fact that both spend and income provide vital role in clustering customers.

2

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

$$x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad A = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \quad (18)$$

$$f\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^T \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (19)$$

$$f\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) = [\alpha \quad \beta] \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (20)$$

$$f\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) = [\alpha a + \beta c \quad \alpha c + \beta b] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (21)$$

$$f\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right) = [\alpha^2 a + 2\alpha\beta c + \beta^2 b] \quad (22)$$

$$f(x) = [\alpha^2 a + 2\alpha\beta c + \beta^2 b] \quad (23)$$

$$\frac{df}{dx} f(x) = \frac{df}{dx} [\alpha^2 a + 2\alpha\beta c + \beta^2 b] \quad \frac{df}{dx} = \left[\frac{df}{d\alpha} \right] \quad (24)$$

$$\frac{df}{dx} f(x) = \left[\frac{df}{d\alpha} \right] [\alpha^2 a + 2\alpha\beta c + \beta^2 b] \quad (25)$$

$$\frac{df}{dx} f(x) = \left[\frac{df}{d\alpha} \alpha^2 a + 2\alpha\beta c + \beta^2 b \right] \quad (26)$$

$$\frac{df}{dx} f(x) = \begin{bmatrix} 2\alpha a + 2\beta c \\ \alpha c + 2\beta b \end{bmatrix} \quad (27)$$

$$2Ax = \begin{bmatrix} 2\alpha a + 2\beta c \\ \alpha c + 2\beta b \end{bmatrix} \quad (28)$$

$$2 \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 2\alpha a + 2\beta c \\ \alpha c + 2\beta b \end{bmatrix} \quad (29)$$

$$\begin{bmatrix} 2\alpha a + 2\beta c \\ \alpha c + 2\beta b \end{bmatrix} = \begin{bmatrix} 2\alpha a + 2\beta c \\ \alpha c + 2\beta b \end{bmatrix} \quad (30)$$

$$(31)$$

3

The main steps for each iteration are 1) finding the distances between the center and each point and 2) updating the location of the centroids

4

Sk.learn uses euclidean distance or L_2 norm

5

Given the fact that that maximum amount of clusters is a finite number which is equivalent the amount of points therefore it must converge.

6

If the position of the centers are randomly initialized then you can have random sets of cluster for the same data and therefore the results are random

7

When there the same number of clusters as data points in our data

Part II