zl7455

Assignment 8 - ECE 4200

$\{$ Name: Zihao Liu
NetID: zl7455 $\}$

Problem #1:

$$w \in \mathbb{R}^{k \times d} = \left.\begin{array}{c} \\ k \\ \\ \end{array}\right\{ \underbrace{\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1d} \\ w_{21} & & & \\ w_{31} & & \ddots & \\ \vdots & & & \\ w_{k1} & & & w_{kd} \end{bmatrix}}_{d}$$

$$w_{ij} \in \left\{-\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}\right\}$$

$$P\left(w_{ij} = \frac{-1}{\sqrt{k}}\right) = P\left(w_{ij} = \frac{+1}{\sqrt{k}}\right) = \frac{1}{2}$$

1. let $\vec{x} \in \mathbb{R}^d$, show $\mathbb{E}\left(\|w\vec{x}\|_2^2\right) = \|\vec{x}\|_2^2$

$\vec{x} = (x^1 \ldots x^d)$

$$w\vec{x} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1d} \\ \vdots & & & \\ w_{k1} & \cdots & & w_{kd} \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^d \end{bmatrix}$$

$$= \begin{bmatrix} w_{11}x_1 + w_{12}x_2 + \cdots + w_{1d}x_d \\ \vdots \\ w_{k1}x_1 + \cdots + w_{kd}x_d \end{bmatrix}$$

$$\|w\vec{x}\|_2^2 = (w_{11}x_1 + \cdots + w_{1d}x_d)^2 + \cdots + (w_{k1}x_1 + \cdots + w_{kd}x_d)^2$$

$$\mathbb{E}\left(\|w\vec{x}\|_2^2\right) = \mathbb{E}\left[\sum_{j=1}^{k} (w_{j1}x_1 + \cdots + w_{jd}x_d)^2\right]$$

(continued next page)

71755

$$\mathbb{E}(\|Wx\|_2^2) = \mathbb{E}\left((w_{11}x_1 + \dots + w_{1d}x_d)^2 + \dots + (w_{k1}x_1 + \dots + w_{kd}x_d)^2 \right)$$

$$= \mathbb{E}\left((w_{11}x_1 + \dots + w_{1d}x_d)^2\right) + \dots + \mathbb{E}\left((w_{k1}x_1 + \dots + w_{kd}x_d)^2\right)$$

linearity of expected values

$$\Rightarrow \boxed{\mathbb{E}(w_{ij}^2) = \frac{1}{k}.} \quad b/c \quad \left\{ \begin{array}{l} w_{ij} \in \{-1/\sqrt{k}, 1/\sqrt{k}\} \\ w_{ij}^2 \in \{\frac{1}{k}\} \end{array} \right\}$$

$$= \mathbb{E}\left(w_{11}^2 x_1^2 + w_{12}^2 x_2^2 + \dots + \cancel{w_{11}w_{12}x_1x_2 + \dots}\xrightarrow{}0\right) + \dots \mathbb{E}\left(w_{k1}^2x_1^2 + \dots w_{kd}^2 x_d^2 + \cancel{w_{j1}w_{j2}x_1x_2 + \dots}\xrightarrow{}0\right)$$

$$= \mathbb{E}(w_{11}^2 x_1^2) + \mathbb{E}(w_{12}^2 x_2^2) + \dots + \mathbb{E}(w_{k1}^2 x_1^2) + \dots \mathbb{E}(w_{kd}^2 x_d^2)$$

$$= \sum_{j=1}^{k} \cancel{\mathbb{E}(w_{ij})^2}\xrightarrow{}1/k \; \|\vec{x}\|_2^2 = \sum_{j=1}^{k} 1/k \, \|\vec{x}\|_2^2$$

$$= \boxed{\|\vec{x}\|_2^2}$$

$\therefore$ QED

2. This random matrix ($w_{ij} \in \{-1/\sqrt{k}, \frac{1}{\sqrt{k}}\}$) 71755

might be more computationally efficient as it only has two outcomes and might be better for storage/memory. The gaussian distribution is more complex involving exponential functions, which may be more computationally intensive.

Problem #2 (Clustering):

7/7/15

Clustering algorithm:

① Init: choose k data points as initial $\{\bar{c}_1, \dots, \bar{c}_k\}$.

② Repeat:

- assign each $\bar{x}_i$ to the closest $\bar{c}_j$ to get clusters.
- change $\{\bar{c}_1 \dots \bar{c}_k\}$ to the new cluster means

③ End: if no new cluster assignment changes

Given a cluster $C = \{\bar{x}_1, \dots \bar{x}_{|C|}\}$

Best $\bar{c}$ is:

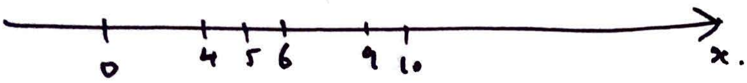
$$\bar{c} = \frac{1}{|C|} \sum_{1 \le i \le |C|} \bar{x}_i$$

optimal solution

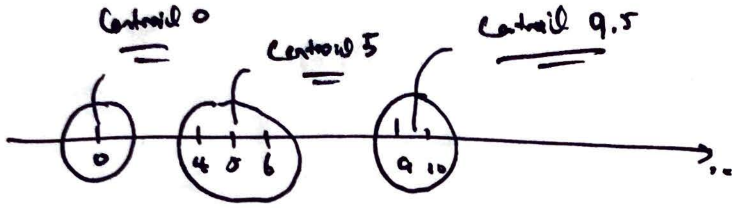
We have 6 real numbers, and initial 3 distinct cluster centers

7/7/55

numbers: $S = \{0, 4, 5, 6, 9, 10\}$



The best clustering Sol'n is (3 clusters):



Initialization of Centroids:

$$C = \{c_1^i, c_2^i, c_3^i\} = \{4, 9, 10\}$$

Iteration 1:

Cluster 1: $\{0, 4, 5, 6\}$

Cluster 2: $\{9\}$ no closer to 9

Cluster 3: $\{10\}$ m closest to 10

New Centroid:

cluster 1: $\frac{0+4+5+6}{4} = 15/4 = \underline{3.75}$

cluster 2: $9/1 = \underline{9}$

cluster 3: $10/1 = \underline{10}$

Iteration 2:

7175

Cluster 1: {0,4,5,6}

Cluster 2: {9}

Cluster 3: {10}

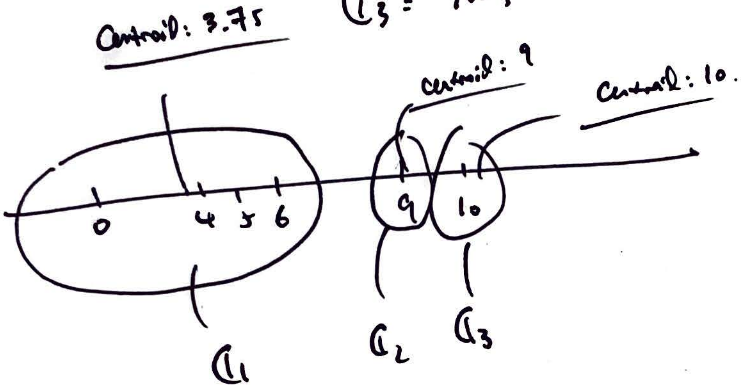
No change to cluster assignment

Final Cluster:

$C_1 = \{0,4,5,6\}$

$C_2 = \{9\}$

$C_3 = \{10\}$



∴ This is suboptimal. Not the most efficient way of clustering set S. Optimal solution is shown earlier.

71755

Problem 3:

$$\sum_{x_i\in C} \|x_i - c\|_2^2 = \sum_{x_i \in C} \|x_i - c_{av} + c_{av} - c\|^2$$

$$= \sum_{x_i\in C} \|x_i - c_{av}\|^2 + 2(c_{av}-c)\cdot\sum_{x_i\in C}(x_i - c_{av}) + |C|\cdot\|c_{av}-c\|^2$$

cardinality of set C.

Since c_{av} is centroid:

$$c_{av} = \frac{1}{|C|}\sum_{x_i\in C} x_i \ , \quad \sum_{x_i\in C}(x_i - c_{av}) = \sum_{x_i\in C} x_i - \sum_{x_i\in C} c_{av}$$

$$= \sum_{x_i\in C} x_i - \sum_{x_i\in C}\left[\frac{1}{|C|}\sum_{x_i\in C} x_i\right]$$

$$= \sum_{x_i\in C} x_i - \cancel{|C|}\cdot\frac{1}{\cancel{|C|}}\sum_{x_i\in C} x_i$$

$$= \sum_{x_i\in C} x_i - \sum_{x_i\in C} x_i = \boxed{0}$$

$$\Rightarrow \sum_{x_i\in C} \|x_i - c\|_2^2 = \boxed{\sum_{x_i\in C}\|x_i - c_{av}\|^2 + |C|\,\|c_{av}-c\|^2}$$

Therefore,

$$\Rightarrow \sum_{x_i\in C}\|x_i - c_{av}\|^2 + \underbrace{\overbrace{|C|\,\|c_{av}-c\|^2}^{\geq 0}}_{\text{Additional term}} \geq \sum_{x_i\in C}\|x_i - c_{av}\|_2^2$$

$\therefore$ QED.