Student Name: Minh Phan
Course: AMATH 352

Homework 1

---

1 Reading

1. Defining the clustering problem:
   The clustering problem proposes a question whether we can "cluster" or categorize vectors/items into different groups so that vectors/items in the same group have similar features. In terms of vectors in a vector space, we consider it as the task of grouping existing vectors into clusters of vectors that are close to each other.

2. Defining the k-cluster algorithm:
   The k-cluster algorithm is one of the best algorithms used for clustering problems. It has two main steps, iterated in a loop until we reach some sort of value representing the ideal "closeness" (or how similar can the features has to be in a set of items) and then stop. Therefore, this approach is convergent. Given a list of N vectors $x_1, x_2, ..., x_N$, the two main steps are:

   (a) Assign each vector x_i into the group associated with the nearest representative. The k representatives can be chosen randomly at first and will be modified as the loop goes on.

   (b) Update the k representatives by setting z_j to be the mean of the vectors in group j. We will then evaluate whether these updated representatives can be used to cluster the vectors. If not, return to step 1.

3. The intuition of the k-cluster algorithm:

   We can use the k-cluster algorithm based on the intuition that we can repeatedly reposition the centroids of the clusters after each iteration using distance functions (norms are some of them) to calculate the best centroid positions. By shifting around, these centroids will eventually be able to place themselves into appropriate positions that would represent the clusters.

2 Exercises

1. Question 1

   (a)
$$LHS = ||\underline{x} + \underline{y}||_2^2 = ((x_1 + y_1)^2 + (x_2 + y_2)^2 + ... + (x_d + y_d)^2))^{(1/2)\cdot 2}$$
$$= (x_1^2 + x_2^2 + ... + x_d^2) + (y_1^2 + y_2^2 + ... + y_d^2) + (2x_1y_1 + 2x_2y_2 + ... + 2x_dy_d)$$
$$= (x_1^2 + x_2^2 + ... + x_d^2)^{(1/2)\cdot 2} + (y_1^2 + y_2^2 + ... + y_d^2)^{(1/2)\cdot 2)} + 2(x_1y_1 + x_2y_2 + ... + x_dy_d)$$
$$= ||\underline{x}||_2^2 + ||\underline{y}||_2^2 + 2\underline{x}^T\underline{y} = RHS$$

   (b)
$$LHS = ||\underline{x} - \underline{y}||_2^2 + ||\underline{x} - \underline{y}||_2^2 \tag{1}$$

   From part a we already proved that

$$||\underline{x} + \underline{y}||_2^2 = ||\underline{x}||_2^2 + ||\underline{y}||_2^2 + 2\underline{x}^T\underline{y} \tag{2}$$

Using the same approach, we can also prove that:

$$||\underline{x}-\underline{y}||_2^2 = ((x_1-y_1)^2+(x_2-y_2)^2+...+(x_d-y_d)^2))^{(1/2)\cdot 2}$$

$$= (x_1^2+x_2^2+...+x_d^2)+(y_1^2+y_2^2+...+y_d^2)-(2x_1y_1+2x_2y_2+...+2x_dy_d)$$

$$= (x_1^2+x_2^2+...+x_d^2)^{(1/2)\cdot 2}+(y_1^2+y_2^2+...+y_d^2)^{(1/2)\cdot 2)}-2(x_1y_1+x_2y_2+...+x_dy_d)$$

$$= ||\underline{x}||_2^2+||\underline{y}||_2^2-2\underline{x}^T\underline{y} \tag{3}$$

From (1), (2) and (3):

$$LHS = ||\underline{x}-\underline{y}||_2^2+||\underline{x}-\underline{y}||_2^2 = ||\underline{x}||_2^2+||\underline{y}||_2^2-2\underline{x}^T\underline{y}+||\underline{x}||_2^2+||\underline{y}||_2^2+2\underline{x}^T\underline{y}$$

$$= 2||\underline{x}||_2^2+2||\underline{y}||_2^2 = RHS$$

2. Question 2

   (a) Using the four axioms to prove that this identity is a norm:

   - (Non negativity) We have both $|x_1+x_2| \geq 0$ and $|x_1-x_2| \geq 0$ for any value of x_1 and $x_2 \in \mathbb{R}^2$ therefore $\max(|x_1+x_2|, |x_1-x_2|) \geq 0$ hence $||\underline{x}|| \geq 0$
   - (Definiteness) Suppose that $||x|| = 0$ and either $x_1+x_2 \neq 0$ or $x_1-x_2 \neq 0$, then $\max(|x_1+x_2|, |x_1-x_2|) > 0$, which leads to contradiction
   - (Positive homogeneity) $||\alpha\underline{x}|| = \max(|\alpha(x_1+x_2)|, |\alpha(x_1-x_2)|)$
   $= |\alpha|\max(|x_1+x_2|, |x_1-x_2|)$ (because the maximum value is guaranteed to be positive, α must also be in absolute to make sure the value stays positive)
   $= |\alpha| \cdot ||\underline{x}||$
   - (Triangle Inequality) $||\underline{x}+\underline{y}|| = \max(|x_1+x_2+y_1+y_2|, |x_1-x_2+y_1-y_2|)$
   We know that for scalars that $|a+b| < |a|+|b|$ for any scalar $a, b \in \mathbb{R}$.
   Let $a = |x_1+x_2|$ and $b = |y_1+y_2|$ then $|x_1+x_2+y_1+y_2| < |x_1+x_2|+|y_1+y_2|$. The same goes with $|x_1-x_2+y_1-y_2| < |x_1-x_2|+|y_1-y_2|$
   Therefore:
   $\max(|x_1+x_2+y_1+y_2|, |x_1-x_2+y_1-y_2|) < \max(|x_1+x_2|+|y_1+y_2|, |x_1-x_2|+|y_1-y_2|) \leq \max(|x_1+x_2|, |x_1-x_2|)+\max(|y_1+y_2|, |y_1-y_2|) = ||\underline{x}||+||\underline{y}||$

   (b) Using the four axioms to prove that this identity is a norm:

   - (Non negativity) The 2-norm is defined as $||\underline{x}||_2 := \sqrt{x_1^2+x_2^2+...+x_d^2} \geq 0$ for all $x \in \mathbb{R}^2$
   - (Definiteness) Suppose $||\underline{x}||_2 = 0$ and $x_k \neq 0$ (with $k \leq d$) then $\sqrt{x_1^2+x_2^2+...+x_d^2} = ||\underline{x}||_2 \geq 0$, which leads to contradiction
   - (Positive homogeneity) $||\alpha\underline{x}||_2 = \sqrt{\sum_{i=1}^d(\alpha x_i)^2} = \sqrt{\sum_{i=1}^d \alpha^2 x_i^2} = \sqrt{\alpha^2\sum_{i=1}^d x_i^2} = |\alpha|\sqrt{\sum_{i=1}^d x_i^2}$
   $= |\alpha| \cdot ||\underline{x}||_2$
   - (Triangle Inequality) From question 1a, we have proved that $||\underline{x}+\underline{y}||_2^2 = ||\underline{x}||_2^2+2\underline{x}^T\underline{y}+||\underline{y}||_2^2$
   $\leq ||\underline{x}||_2^2+2|\underline{x}^T\underline{y}|+||\underline{y}||_2^2 \leq ||\underline{x}||_2^2+2||\underline{x}||||\underline{y}||+||\underline{y}||_2^2 = (||\underline{x}||+||\underline{y}||)^2$
   Take the square root the two sides we have: $||\underline{x}+\underline{y}|| \leq ||\underline{x}||+||\underline{y}||$

   (c) We know that $||\underline{x}||_\infty := \max(|x_j|)$ with $j = 1, 2, ..., d$. Using the four axioms to prove that this identity is a norm

   - (Non negativity) Because every single element $|x_k|$ with $k = 1, 2, ...d)$ is non-negative, therefore $\max(|x1|, |x2|, ..., |x_d|) = ||\underline{x}||_\infty \geq 0$ with all $\underline{x} \in \mathbb{R}^d$
   - (Definiteness) Suppose $||\underline{x}||_\infty = 0$ and $x_k \neq 0$ (with $k \leq d$) then $\max(|x1|, |x2|, ..., |x_d|) = ||\underline{x}||_\infty > 0$ hence leading to contradiction

- (Positive homogeneity) $||\alpha\underline{x}||_\infty = \max(|\alpha x_j|)$ with j from 1 to d $= |\alpha| \max(|x_j|)$ (to ensure that the value stays positive) $= |\alpha| \cdot ||\underline{x}||_\infty$
- (Triangle Inequality) $||\underline{x} + \underline{y}||_\infty = \max(|x_j + y_j|) \leq \max(|x_j| + |y_j|)$ $\leq \max(|x_j|) + \max(|y_j|) = ||\underline{x}||_\infty + ||\underline{y}||_\infty$ for j values from 1 to d

3. Question 3

- $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

  The angle is: $\arccos(\frac{1(-1)+0\cdot 0}{\sqrt{1^2+0^2}\sqrt{(-1)^2+0^2}}) = \arccos(-1)$
- $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1/2 \end{bmatrix}$

  The angle is $\arccos(\frac{1\cdot 1+0\cdot 1+0\cdot 1/2}{\sqrt{1^2+0^2+0^2}\sqrt{1^2+1^2+(1/2)^2}}) = \arccos(2/3)$
- $\begin{bmatrix} 1 \\ \sqrt{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1/2 \\ 2 \end{bmatrix}$

  The angle is $\arccos(\frac{1(-1)+\sqrt{3}\cdot 0+1(1/2)+0\cdot 2}{\sqrt{1^2+(\sqrt{3})^2+1^2+0^2)(\sqrt{(-1)^2+0^2+(1/2)^2+2^2}}}) = \arccos(\frac{-1}{\sqrt{105}})$