HW 01 Pavel Stepanov

1.

$$x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{matrix} x_1 \\ y_1 \end{matrix} \qquad x_2 = \begin{bmatrix} 10 \\ 18 \end{bmatrix} \begin{matrix} x_2 \\ y_2 \end{matrix}$$

Distance formula : $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

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

$$d = \sqrt{9^2 + 16^2} = \sqrt{337} = 18.357$$

(1) Based on L2 norm (Euclideum norm)

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ by } \|X\|_2 = \sqrt{\sum_{k=1}^{n} |X_k|^2}$$

So we have for:

$$X_1 = \|X_1\|_2 = \sqrt{|1|^2 + |2|^2} = \sqrt{3}$$

$$X_2 = \|X_2\|_2 = \sqrt{|10|^2 + |18|^2} =$$

$$\Rightarrow = \sqrt{|100| + |324|} = 2\sqrt{106}$$

and we have: $x_2 - x_1 = 2\sqrt{106} - \sqrt{3} = 18.859$

(2) Based on the L1 norm:

$$\|X\|_1 = \sum |x_k|^1$$

So we have for:

$$x_1 = \|X_1\|_1 = |1|^1 + |2|^1 = 3$$

$$x_1 = \|X_2\|_1 = |10|^1 + |18|^1 = 28$$

and we have: $x_2 - x_1 = 28 - 3 = 25$

$$* \quad \|X\|_p = \left(\sum |x_k|^p\right)^{1/p}$$

(3) Based on the $L_\infty$ Norm:

$$\|X\|_\infty = \max_i |x_i|$$

So we have for:

$$x_1 = \|x_1\|_\infty = \max(|1|, |2|)$$
$$\Rightarrow \max(1, 2) = 2$$

$$x_2 = \|x_2\|_\infty = \max(|10|, |18|)$$
$$\Rightarrow \max(10, 18) = 18$$

So, $X_2 - X_1 = 18 - 2 = 16$

* Assume that we use a clustering method similar to k-mean, and this method could use any type of vector norms as distance measure. Then, does the L∞ norm-based distance measure make sense for this application?

- The $L_\infty$ norm, which considers the max absolute difference between any feature might not be the best choice In the context because It Is Only accounts for the feature with largest difference Ignores other.
- If Income and spend differ significantly $L_\infty$ norm will only reflect the maximum of these two difference, which might not capture the Overall dissimilarity customers based on the both features

So we have, that $L_\infty$ norm measure does not make sense for this application If we want a comprehensive understanding Of how customers differ In both Income and Spend.

$\boxed{2.}$ Scalar valued function of a vector variable: $f(x) = x^T A x$

We have two elements: $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, and $A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$

$$\frac{df}{dx} = \begin{bmatrix} \frac{df}{d\alpha} \\ \frac{df}{d\beta} \end{bmatrix}, \quad \text{Show that } \frac{df}{dx} = 2Ax$$

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

we know that: $f(x) = x^T A x$, $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

$x^T = [\alpha, \beta]$

S.1. $f(x) = x^T Ax$

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

- $x^T Ax = [\alpha, \beta] \begin{bmatrix} a\alpha + c\beta \\ c\alpha + b\beta \end{bmatrix}$

$\Rightarrow \alpha(a\alpha + c\beta) + \beta(c\alpha + b\beta)$

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

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

$$f(x) = a\alpha^2 + 2\alpha\beta c + b\beta^2$$

5.2. $$2Ax = 2 \cdot \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

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

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

5.3 a) $\frac{df}{d\alpha}$ of $f(x) = a\alpha^2 + 2\alpha\beta c + b\beta^2$

$$\frac{df}{d\alpha} = \frac{d}{d\alpha}(a\alpha^2 + 2\alpha\beta c + b\beta^2)$$

$$= 2a\alpha + 2\beta c + \cancel{b\beta^2}$$

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

b) $\frac{df}{d\beta}$ of $f(x) = a\alpha^2 + 2\alpha\beta c + b\beta^2$

$$\frac{df}{d\beta} = \frac{d}{d\beta}(a\alpha^2 + 2\alpha\beta c + b\beta^2)$$

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

Now we have that the gradient of

$$\frac{df}{dx} = \begin{bmatrix} \frac{df}{d\alpha} \\ \frac{df}{d\beta} \end{bmatrix} = \begin{bmatrix} 2a\alpha + 2c\beta \\ 2b\beta + 2\alpha c \end{bmatrix}$$

And we can make a conclusion that,

yes $\frac{df}{dx}$ is exactly the same as $2Ax$

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

We showed that derivative of function $f(x) = x^T Ax$ with respect to the vector $x$ is Indeed $\frac{df}{dx} = 2Ax$