Mathematics of Data Sceience

Tutorial, October 13

Hung-Hsu Chou

Mathematics of Information Processing, RWTH

Winter Semester 2021/2022

Outline

Outline

Logistics

- Tutorials will not be recorded. Slides will be available on Moodle.

Logistics

- Tutorials will not be recorded. Slides will be available on Moodle.
- Exercise sheets will be posted every Wednesday. First one is on October 13.

Logistics

- Tutorials will not be recorded. Slides will be available on Moodle.
- Exercise sheets will be posted every Wednesday. First one is on October 13.
- The solutions will be covered during the tutorials a week after the sheets are posted.

Logistics

- Tutorials will not be recorded. Slides will be available on Moodle.
- Exercise sheets will be posted every Wednesday. First one is on October 13.
- The solutions will be covered during the tutorials a week after the sheets are posted.
- Dynexite Test takes places every two weeks. First one opens from October 27 (after tutorial) - October 31. Duration is 30 - 45 minutes after start.

Outline

Quote

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.

- Carl Friedrich Gauss

Outline

Decision Making

Measure $\rightarrow$ Compare $\rightarrow$ Make Decision

Decision Making

Measure $\rightarrow$ Compare $\rightarrow$ Make Decision

Example

You tried to call your friend at 1AM, 2AM, and 3AM. He/she was happy, quiet, and angry, respectively. It is 4AM now. Should you call?

Decision Making

Measure $\rightarrow$ Compare $\rightarrow$ Make Decision

Example

You tried to call your friend at 1AM, 2AM, and 3AM. He/she was happy, quiet, and angry, respectively. It is 4AM now. Should you call?

This is a dangerous experiment and is not recommended!

Example

Question

1. *Let* x, y, z *be wallets with* $5, 10, 20$ *Euros. Which one would you pick?*

Example

Question

1. *Let* x, y, z *be wallets with* $5, 10, 20$ *Euros. Which one would you pick?*
2. *Let* x, y, z *be candidates of your project teammates. Which one would you pick? (The score is out of 10.)*

	Intelligence	*Friendliness*	*Punctuality (On Time)*
x	9	5	5
y	6	8	6
z	7	7	7

Table: *Choose carefully!*

Example

Question

1. *Let* x, y, z *be wallets with* $5, 10, 20$ *Euros. Which one would you pick?*
2. *Let* x, y, z *be candidates of your project teammates. Which one would you pick? (The score is out of 10.)*

	Intelligence	*Friendliness*	*Punctuality (On Time)*
x	*9*	*5*	*5*
y	*6*	*8*	*6*
z	*7*	*7*	*7*

Table: *Choose carefully!*

People might agree on the answer for the first question but not the second, because the lack of order. How should we proceed?

Map for Order

Mathematically we can treat x, y, z as vectors in $\mathbb{R}^3$, or $\mathbb{R}^n$ in general, i.e.

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

Because the only thing we can compare/order is the real number, we need to create a map $f : \mathbb{R}^3 \to \mathbb{R}$ so that we can compare/order $f(x), f(y), f(z)$. One can think of those as the score of x, y, z.

Outline

Norm

One common class of mapping is called **NORM**. The common examples are the ℓ_p norm defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

for all $1 \leq p < \infty$. The infinity norm is defined as

$$\|x\|_\infty = \max_i |x_i|.$$

Decision with Norms

Different norms might gives different results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $f(x) = \|x\|_1$,

Decision with Norms

Different norms might gives different results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $f(x) = \|x\|_1$, then $f(x) < f(y) < f(z)$.
2. If $f(x) = \|x\|_2$,

Decision with Norms

Different norms might gives different results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $f(x) = \|x\|_1$, then $f(x) < f(y) < f(z)$.
2. If $f(x) = \|x\|_2$, then $f(x) < f(y) < f(z)$.
3. If $f(x) = \|x\|_\infty$,

Decision with Norms

Different norms might gives different results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $f(x) = \|x\|_1$, then $f(x) < f(y) < f(z)$.
2. If $f(x) = \|x\|_2$, then $f(x) < f(y) < f(z)$.
3. If $f(x) = \|x\|_\infty$, then $f(z) < f(y) < f(x)$.

Different norms are useful in different circumstances.

Outline

Linear Map

Another common class of mapping is called **LINEAR MAP**. The map takes the form

$$f(x) = \sum_{i=1}^{n} w_i x_i$$

for $w_i \in \mathbb{R}$ (or $\mathbb{C}$ but we will stick with $\mathbb{R}$ for now). The parameter w is also sometimes known as the weight, corresponds to how we value each entry of x.

Decision with Linear Maps

Similar to norms, the choice of w affects the results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $w = [1, 1, 1]$,

Decision with Linear Maps

Similar to norms, the choice of w affects the results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $w = [1, 1, 1]$, then $f(x) < f(y) < f(z)$.
2. If $w = [3, 2, 1]$,

Decision with Linear Maps

Similar to norms, the choice of w affects the results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $w = [1, 1, 1]$, then $f(x) < f(y) < f(z)$.
2. If $w = [3, 2, 1]$, then $f(y) < f(x) = f(z)$.
3. If $w = [0, 2, 0]$,

Decision with Linear Maps

Similar to norms, the choice of w affects the results. Recall that

$$x = [9, 5, 5], \quad y = [6, 8, 6], \quad z = [7, 7, 7].$$

1. If $w = [1, 1, 1]$, then $f(x) < f(y) < f(z)$.
2. If $w = [3, 2, 1]$, then $f(y) < f(x) = f(z)$.
3. If $w = [0, 2, 0]$, then $f(x) < f(z) < f(y)$.

In fact, there is a compact way to write the linear map using matrices.

Matrices

Suppose we decide to use the linear map with $w = [3, 2, 1]$. To compute the result we can simply compute Aw, where

$$A = \begin{bmatrix} — & \mathrm{x} & — \\ — & \mathrm{y} & — \\ — & \mathrm{z} & — \end{bmatrix} = \begin{bmatrix} 9 & 5 & 5 \\ 6 & 8 & 6 \\ 7 & 7 & 7 \end{bmatrix}$$

and

$$Aw = \begin{bmatrix} 9 & 5 & 5 \\ 6 & 8 & 6 \\ 7 & 7 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 42 & (\text{Score of } x) \\ 40 & (\text{Score of } y) \\ 42 & (\text{Score of } z) \end{bmatrix}.$$

It is easy to see that x and z have the best score.

Outline

Computation Time

Sometimes we need to multiply matrices, but the computation time might explodes if we do it directly.

Question

1. *Let* $A = \begin{bmatrix} 9 & 5 & 5 \\ 6 & 8 & 6 \\ 7 & 7 & 7 \end{bmatrix}$. *What is* A^4*?*

Computation Time

Sometimes we need to multiply matrices, but the computation time might explodes if we do it directly.

Question

1. *Let* $A = \begin{bmatrix} 9 & 5 & 5 \\ 6 & 8 & 6 \\ 7 & 7 & 7 \end{bmatrix}$. *What is* A^4*?*
2. *Let* $A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 7 \end{bmatrix}$. *What is* A^4*?*

Computation Time

Sometimes we need to multiply matrices, but the computation time might explodes if we do it directly.

Question

1. *Let* $A = \begin{bmatrix} 9 & 5 & 5 \\ 6 & 8 & 6 \\ 7 & 7 & 7 \end{bmatrix}$. *What is* A^4*?*
2. *Let* $A = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 7 \end{bmatrix}$. *What is* A^4*?*

Diagonal matrices are much easier work with!

Eigenvector

For any $A \in \mathbb{R}^{n\times n}$, if we can express it as $A = VDV^*$, where D is diagonal and V is unitary, i.e. $VV^* = V^*V = I$, then

Eigenvector

For any $A \in \mathbb{R}^{n \times n}$, if we can express it as $A = VDV^*$, where D is diagonal and V is unitary, i.e. $VV^* = V^*V = I$, then

$$
\begin{aligned}
A^k &= \underbrace{(VDV^*)\cdots(VDV^*)}_{k \text{ times}} \\
&= VD\underbrace{(V^*VD)\cdots(V^*VD)}_{(k-1) \text{ times}}V^* \\
&= VD\underbrace{D\cdots D}_{(k-1) \text{ times}}V^* \\
&= VD^kV^*
\end{aligned}
$$

which is easy to compute because we only need to compute D^k.