

The principle of Mja

Wdataorg

December 18, 2022

Contents

1	Add a sample	1
1.1	Input sample	1
1.2	Generate sample group information	1
1.3	Why do we need least 5 samples?	2
2	Calculate with formula	3
2.1	Cosine similarity	3
2.2	A example	3
2.3	Sort cosine similarity list	3
2.4	Delete samples with low similarity	3
2.5	Regression	4

1 Add a sample

First, Mja need least 5 samples to predict

In Mja Software, we use **Add a sample** button to add a characteristic information, after input this information, Mja will convert them into mathematical models

Like this page.

1.1 Input sample

When you have had a exam or test, and you know the score of this exam, you can use **Add a sample** button to add a information, plenty of information will make predictions more accurate

If user input all information in this picture, they can click this button, Mja will generate a five dimensional model, and save it into a file

1.2 Generate sample group information

Mja will convert these survey data into a model.How does Mja convert?

Mja
Add a sample

What do you think about the difficult of this exam/test
Do you review exam content before you have this exam/test
Is it a exam or test
How are you feeling before the exam/test
Is this a sudden test
Please input your score.
Please enter the full score of this test/exam
Submit sample

Mja will convert each item of data in the survey into a coordinate of the five dimensional model according to the following table

Model generation table				
Difficult			Have Reviewed	
Easy	Normal	Hard	Have	Haven't
1	2	3	1	2
Feeling				
Very good	Good	Normal	Bad	Very Bad
1	2	3	4	5
Is a Sudden Test				
Yes		-	No	
1		-	2	

Such as, user input Normal, Have Reviewed, Good, Yes, Mja will generate:
[2, 1, 2, 1]

The last dimension data is the score, where the score refers to a ratio. For example, if the user enters 150 in the full score and 90 in the score, the one-dimensional data will be $\frac{90}{150} \times 100$, so it is 60

1.3 Why do we need least 5 samples?

When you have no least 5 example, you can not use Predict Grades Program, Because Mja Extract the features of the first \sqrt{n} samples, When you have only 3 samples, it will be $\sqrt{3} \approx 1$, but Mja uses KNN, KNN need least 2 samples, so $n \geq 4$, Because $2^2 = 4$, But to make the results accurate, so Mja need least 5 samples

2 Calculate with formula

When we have enough samples, we can predict grades, predict grades use cosine similarity, It can calculate the cosine of the angle between two coordinates, The closer the cosine value is to 1, the closer the two vectors are.

Mja used cosine similarity to predict grades.

2.1 Cosine similarity

The formula of cosine similarity like this:

$$\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (1)$$

2.2 A example

I will use a example to tell reader how to use this formula
such as we have this data:

Difficult	Have Reviewed	Feeling	Is a Sudden Test	Score	Full Score
Normal	Have	Good	Yes	100	150
2	1	2	1	67	-

and we have this new list

3	1	2	2	70	-
---	---	---	---	----	---

Substitute into this formula

$$\cos(\theta) = \frac{2 \times 3 + 1 \times 1 + 2 \times 2 + 1 \times 2}{\sqrt{2^2 + 1^2 + 2^2 + 1^2} \times \sqrt{3^2 + 1^2 + 2^2 + 2^2}} \approx 0.97 \quad (2)$$

So we can they these samples are very close

2.3 Sort cosine similarity list

First, we should use this formula to calculate the similarity of samples to be tested and all known samples. And then sort them

For example, we have similarity list: [(0.998, 90), (0.678, 80), (0.892, 95), (0.977, 100)]

After sort them, we can get the new similarity list: [(0.998, 90), (0.977, 100), (0.892, 95), (0.678, 80)]

2.4 Delete samples with low similarity

Then, we have a sorted list.

We only need first $\sqrt{\text{list.length}()}$ elements

Such as, similarity list: [(0.998, 90), (0.977, 100), (0.892, 95), (0.678, 80)]

$\sqrt{\text{list.length}()} = 2$, so we only need [(0.998, 90), (0.977, 100)]

2.5 Regression

At last, we should **Regression** similarity list.

This step only requires the average similarity value

The value of similarity list $[(0.998, 90), (0.977, 1000)]$ is $[90, 100]$, average it, we can get the results 95.

Congratulations, you successfully completed the result forecast!