ZHONG Qiaoyang

24112456g

Assignment 1

Q1

Task 1

Scale-up: make the single machine bigger (like more GPUs, CPUs, bigger RAM, VRAM), so that we can handle more resources, like run the model with larger parameters.

Scale-out: distribute workload across multiple machines, like MOE architecture supports multi-clusters training.

Task 2
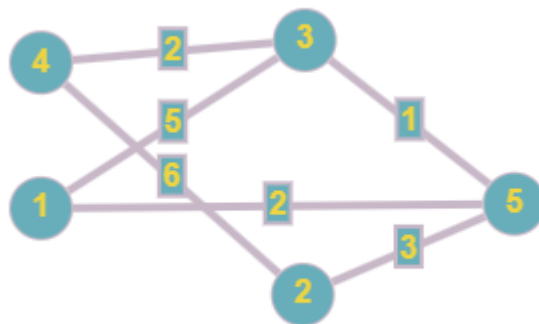
Spatial: space in data (like image coordinate)

Temporal: time in data (like time series in video)

Task 3

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

answer is `[0.0, 0.25, 0.125, 1.0, 0.75, 0.375, 0.5, 0.625]`

Task 4

Task 5

Descriptive Analysis: dig out the information in the past

Predictive Analysis: predict the future

Q2

Task 1

```
TF-IDF Matrix:
Document 1: [0.219, 0.439, 0.0, 0.219, 0.28]
Document 2: [0.548, 0.548, 0.0, 0.0, 0.0]
Document 3: [0.0, 0.0, 0.611, 0.548, 0.0]
Document 4: [0.548, 0.0, 0.0, 0.548, 0.0]
Document 5: [0.0, 0.366, 0.815, 0.0, 0.0]
```

Task 2

```
Cosine Similarities of D5 with D1, D2, D3, D4:
Similarity with D1: 0.297
Similarity with D2: 0.290
Similarity with D3: 0.679
Similarity with D4: 0.000
```

D3 is the most similar one to D5.

Q3

Step 1: 1 item

Item	Count	Support
A	4	0.67
B	2	0.33
C	3	0.50
D	4	0.67
E	4	0.67

Step 2: 2 items

- (A, C): Support = 1/6 = 0.17 (Not frequent)
- (A, D): Support = 3/6 = 0.5
- (A, E): Support = 3/6 = 0.5
- (C, D): Support = 2/6 = 0.33 (Not frequent)
- (C, E): Support = 2/6 = 0.33 (Not frequent)
- (D, E): Support = 4/6 = 0.67

Step 3: 3 items

- (A, D, E): T1, T2, T5 → 3; Support = 3/6 = 0.5

Step 4: association rules

- D ⟹ E: 0.67/0.67 = 1.0 (**valid**)
- E ⟹ D: 0.67/0.67 = 1.0 (**valid**)
- (A, D) ⟹ E: 0.5/0.5 = 1.0 (**valid**)
- (A, E) ⟹ D: 0.5/0.5 = 1.0 (**valid**)