



Тест по темам "кластеризация" и "ассоциативные правила". В вопросах на ввод числа, ответ может быть как целочисленным, так и рациональным числом (например 0.1 итп)

В тестовых заданиях первая галочка — правильный ответ, вторая галочка — выбранный ответ. Цвет обозначает, правильно ли в данном пункте поставлена галочка. Если все пункты верные (галочки совпадают / все пункты зеленые), то за задание ставится полный балл, в противном случае ставится 0 баллов.

1. Рассмотрим метод иерархической кластеризации снизу вверх (агломеративная кластеризация) с методом полной связи (complete linkage) в качестве правила вычисления расстояний между кластерами. Обозначим за  $r(i,j)$  - расстояние между кластерами  $i$  и  $j$ . Пусть кластер  $k$  - результат объединения кластеров  $i$  и  $j$ . Тогда расстояние до кластера  $p$  вычисляется как

- ☐ ☐ среднее между  $r(i,p)$  и  $r(j,p)$
- ☒ ☒  $\max(r(i,p), r(j,p))$
- ☐ ☐ взвешенное среднее между  $r(i,p)$  и  $r(j,p)$
- ☐ ☐  $\min(r(i,p), r(j,p))$

**Балл: 2.0**

**Комментарий к правильному ответу:**

2. Пусть  $X, Y$  - подмножества товаров. Поддержка (support) ассоциативного правила  $X \rightarrow Y$  измеряет

- ☐ ☐ долю транзакций, в которых присутствует  $X$
- ☐ ☐ долю транзакций, в которых присутствует  $X$  или  $Y$
- ☐ ☐ долю транзакций, в которых присутствует  $Y$
- ☐ ☐ долю транзакций, в которых присутствует  $Y$ , среди транзакций, содержащих  $X$
- ☐ ☐ долю транзакций, в которых присутствует  $X$ , среди транзакций, содержащих  $Y$
- ☒ ☒ долю транзакций, в которых присутствует  $X$  и  $Y$

**Балл: 2.0**

**Комментарий к правильному ответу:**

3. Дано пять корзин:  $\{A, B, C\}$ ,  $\{A, B\}$ ,  $\{C, D\}$ ,  $\{B\}$ ,  $\{B, C\}$ . Чему равна  $\text{confidence}(\{B\} \rightarrow \{A\})$ ?



Ответ: 0.5

Правильный ответ: 0.5

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

4. Предположим, что покупатели набрали всевозможные сочетания корзин на конечном множестве товаров (при этом все корзины уникальны, встретились по одному разу, есть и одна пустая корзина). Для конкретных товаров A и B чему равно  $\text{confidence}(\{B\} \rightarrow \{A\})$ ?

Ответ: 0.25

Правильный ответ: 0.5

Погрешность: 0.0

Балл: 0

Комментарий к правильному ответу:

5. Рассмотрим случайные события, заключающиеся в присутствии подмножества товаров X и подмножества товаров Y в случайно выбранной транзакции. Допустим эти два события независимы. В этом случае мера lift будет равна

☐ ☐ плюс бесконечности

☒ ☒ единице

☐ ☐ нулю

Балл: 2.0

Комментарий к правильному ответу:

6. Рассмотрим метод иерархической кластеризации снизу вверх (агломеративная кластеризация) с методом средней связи (group average link) в качестве правила вычисления расстояний между кластерами. Обозначим за  $r(i,j)$  - расстояние между кластерами i и j. Пусть кластер k - результат объединения кластеров i и j. Тогда расстояние до кластера p вычисляется как

☐ ☐  $\min(r(i,p), r(j,p))$

☐ ☐ среднее между  $r(i,p)$  и  $r(j,p)$

☒ ☒ взвешенное среднее между  $r(i,p)$  и  $r(j,p)$



☐ ☐  $\max(r(i,p), r(j,p))$

**Балл: 2.0**

**Комментарий к правильному ответу:**

7. Дано пять корзин: {A, B, C}, {A, B}, {C, D}, {B}, {B, C}. Чему равна  $\text{support}(\{B\} \rightarrow \{A\})$ ?

**Ответ: 0.4**

**Правильный ответ: 0.4**

**Погрешность: 0.0**

**Балл: 2.0**

**Комментарий к правильному ответу:**

8. Какая максимальная поддержка (support) может быть у правила с  $\text{confidence}=0.5$ ?

☒ ☒ 0.5

☐ ☐ 0.75

☐ ☐ 0.25

☐ ☐ 1

☐ ☐ 0

**Балл: 0**

**Комментарий к правильному ответу:**