



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. 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: 0.5

Правильный ответ: 0.5

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

2. 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:

3. Рассмотрим метод иерархической кластеризации снизу вверх (агломеративная кластеризация). Какой метод вычисления расстояний между кластерами наименее склонен объединять непохожие кластера на ранних итерациях при наличии между ними узких цепочек из близко расположенных объектов?

☐ ☐ метод средней связи (group average link)

☐ ☐ центроидный метод (pair-group method using the centroid average)

☐ ☒ метод одиночной связи (single linkage)

☒ ☐ метод полной связи (complete linkage)

Балл: 0

Комментарий к правильному ответу:



4. Рассмотрим метод иерархической кластеризации снизу вверх (агломеративная кластеризация) с методом одиночной связи (single 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

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

☐ ☐ среднее между $r(i,p)$ и $r(j,p)$

☒ ☒ $\min(r(i,p), r(j,p))$

☐ ☐ $\max(r(i,p), r(j,p))$

Балл: 2.0

Комментарий к правильному ответу:

5. Сколько ассоциативных правил $X \rightarrow Y$ с непустой причиной (X) и следствием (Y) можно составить из набора, содержащего 10 товаров?

Ответ: 1022

Правильный ответ: 1022.0

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

$2^N - 2$

6. 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

☒ ☒ $\max(r(i,p), r(j,p))$

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

☐ ☐ среднее между $r(i,p)$ и $r(j,p)$



☐ ☐ $\min(r(i,p), r(j,p))$

Балл: 2.0

Комментарий к правильному ответу:

7. Алгоритм Apriori поиска частых наборов товаров без досрочной остановки

☒ ☒ возвращает полный список всех частых наборов товаров

☐ ☐ находит лишь подмножество частых наборов товаров

Балл: 2.0

Комментарий к правильному ответу:

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

☐ ☐ долю транзакций, в которых присутствует Y

☐ ☐ долю транзакций, в которых присутствует X , среди транзакций, содержащих Y

☒ ☒ долю транзакций, в которых присутствует X и Y

☐ ☐ долю транзакций, в которых присутствует X или Y

☐ ☐ долю транзакций, в которых присутствует Y , среди транзакций, содержащих X

☐ ☐ долю транзакций, в которых присутствует X

Балл: 2.0

Комментарий к правильному ответу: