



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. Пусть N -число объектов выборки, I -число итераций метода K -средних. Тогда вычислительная сложность этого метода равна

☐ ☐ $O(N \cdot K \cdot I \cdot I)$

☐ ☐ $O(N \cdot K \cdot K \cdot I)$

☒ ☒ $O(N \cdot K \cdot I)$

☐ ☐ $O(N \cdot N \cdot K \cdot I)$

Балл: 2.0

Комментарий к правильному ответу:

2. Выберите верные утверждения относительно метода DB-scan:

☒ ☒ область отнесения объектов каждому кластеру может быть как выпуклым, так и невыпуклым множеством

☒ ☒ метод имеет возможность фильтрации выбросов, не относя их ни к какому кластеру

☐ ☐ область отнесения объектов каждому кластеру - всегда выпуклый многогранник

☒ ☐ число кластеров определяется автоматически

Балл: 0

Комментарий к правильному ответу:

3. 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 $r(j,p)$

☐ ☒ $\min(r(i,p), r(j,p))$

☐ ☐ среднее между $r(i,p)$ и $r(j,p)$

Балл: 0

Комментарий к правильному ответу:

4. Алгоритм Apriori поиска частых наборов товаров без досрочной остановки

☐ ☐ находит лишь подмножество частых наборов товаров

☒ ☒ возвращает полный список всех частых наборов товаров

Балл: 2.0

Комментарий к правильному ответу:

5. 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:

6. 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

☐ ☐ нулю

☐ ☐ плюс бесконечности

☒ ☒ единице

Балл: 2.0

Комментарий к правильному ответу:

7. 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 полной связи (complete linkage)
- ☐ ☐ центроидный метод (pair-group method using the centroid average)
- ☐ ☒ метод одиночной связи (single linkage)
- ☐ ☐ метод средней связи (group average link)

Балл: 0

Комментарий к правильному ответу:

8. Алгоритм FP-growth 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: