



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

Балл: 2.0

Комментарий к правильному ответу:

2. Выберите верные утверждения относительно метода mean-shift:

- ☒ ☒ область отнесения объектов каждому кластеру может быть как выпуклым, так и невыпуклым множеством
- ☐ ☐ область отнесения объектов каждому кластеру - всегда выпуклый многогранник
- ☒ ☒ число кластеров определяется автоматически
- ☐ ☒ метод имеет возможность фильтрации выбросов, не относя их ни к какому кластеру

Балл: 0

Комментарий к правильному ответу:

3. 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:



4. Ключевая идея алгоритма Apriori по ускоренному поиску частых наборов товаров заключается в

- ☐ ☐ компактном описании списка всех транзакций в виде иерархической структуры
- ☒ ☒ инкрементальном поиске более длинных частых наборов по более коротким частым наборам
- ☐ ☐ инкрементальном поиске более коротких частых наборов по более длинным частым наборам

Балл: 2.0

Комментарий к правильному ответу:

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

- ☐ ☐ $O(N \cdot K \cdot K \cdot I)$
- ☐ ☐ $O(N \cdot K \cdot I \cdot I)$
- ☒ ☒ $O(N \cdot K \cdot I)$
- ☐ ☐ $O(N \cdot N \cdot K \cdot I)$

Балл: 2.0

Комментарий к правильному ответу:

6. Пусть X, Y - подмножества товаров. Уверенность (confidence) ассоциативного правила $X \rightarrow Y$ измеряет

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

Балл: 2.0

Комментарий к правильному ответу:



7. 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

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

☒ ☒ $\max(r(i,p), r(j,p))$

☐ ☐ среднее между $r(i,p)$ и $r(j,p)$

☐ ☐ $\min(r(i,p), r(j,p))$

Балл: 2.0

Комментарий к правильному ответу:

8. Выберите верные утверждения относительно метода DB-scan:

☐ ☐ область отнесения объектов каждому кластеру - всегда выпуклый многогранник

☒ ☒ число кластеров определяется автоматически

☒ ☒ 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

Балл: 2.0

Комментарий к правильному ответу: