



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 задачу ранжирования. Пусть labeled - обучающие выборки, где каждому объекту x поставлена в соответствие вещественная мера релевантности, а preference - обучающие выборки, где для подмножества пар объектов $(x(i), x(j))$ указана информация, какой из объектов пары более релевантен (но неизвестно насколько). Какие классы методов могут обучаться на каких типах обучающих выборок?

- ☒ ☒ поточечные (pointwise) методы - только на labeled , попарные (pairwise) - на preference и на labeled
- ☐ ☐ каждый метод может обучаться и на labeled , и на preference .
- ☐ ☐ поточечные (pointwise) методы - только на preference , попарные (pairwise) - только на labeled
- ☐ ☐ поточечные (pointwise) методы - только на labeled , попарные (pairwise) - только на preference
- ☐ ☐ поточечные (pointwise) методы - на preference и на labeled , попарные (pairwise) - только на labeled

Балл: 2.0

Комментарий к правильному ответу:

2. Pairwise (попарные) подходы более предпочтительны в задаче ранжирования, чем pointwise (поточечные) подходы, поскольку

- ☐ ☐ попарные методы имеют более разумное начальное приближение, позволяющее сойтись за меньшее число итераций
- ☐ ☐ попарные методы не имеют риска сойтись в локальный, а не глобальный оптимум
- ☐ ☐ попарные подходы обучаются быстрее
- ☒ ☒ существует неоднозначность выбора степени релевантности для соблюдения корректного порядка в ранжировании

Балл: 2.0

Комментарий к правильному ответу:



3. Рассмотрим линейную pairwise модель ранжирования $a(x) = w^T x$, обучаемую методом стохастического градиентного спуска с шагом обучения $\eta > 0$, сэмплированием одной пары объектов (i, j) , $i \succ j$ и экспоненциальной функцией потерь $e^{-(a(x_i) - a(x_j))}$. Выберите корректную формулу для обновления весов:

☐ ☐ $w := w + \eta e^{w^T(x_i - x_j)}(x_i - x_j)$

☒ ☒ $w := w + \eta e^{w^T(x_j - x_i)}(x_i - x_j)$

☐ ☐ $w := w + \eta e^{w^T(x_j - x_i)}(x_j - x_i)$

☐ ☐ $w := w + \eta e^{w^T(x_i - x_j)}(x_j - x_i)$

Балл: 2.0

Комментарий к правильному ответу:

4. При расширении выдачи на запрос пользователя в рекомендательной системе, в отличие от информационного поиска (information retrieval), делается предположение о том, что

☒ ☒ пользователь до конца не знает, что он ищет

☐ ☐ пользователь чётко знает, что он ищет

Балл: 2.0

Комментарий к правильному ответу:

5. Недостаток построения рекомендаций с помощью сокращенного сингулярного разложения (pure SVD) матрицы рейтингов заключается в том, что

☐ ☐ хранение результата сокращенного сингулярного разложения требует больше памяти, чем хранение исходной матрицы рейтингов, что создает сложности для большого числа пользователей и товаров

☒ ☒ требуется заполнение пропущенных рейтингов в матрице рейтингов некоторыми числами, которые трактуются наравне с реально полученными рейтингами

☐ ☐ даже небольшое сокращение ранга приводит к полной потере информации об индивидуальных особенностях пользователя с невозможностью в дальнейшем построить для него рекомендацию

Балл: 2.0

Комментарий к правильному ответу:



6. При информационном поиске сформирована выдача из 5 документов, из 2 релевантных документов оба попали в выдачу на позиции 2 и 4. Вычислите меру $AP@5$.

Ответ: 0.5

Правильный ответ: 0.5

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

$$\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

7. Рассмотрим меру Discounted Cumulative Gain (DCG) с функцией расчета выигрыша $Gain(y) = 2^y - 1$ (y -степень релевантности) и функцией дисконтирования за более позднюю выдачу $Discount(i) = 1/i$, где $i=1,2,3,\dots$ - номер позиции. Вычислите $DCG@7$, если релевантными оказались 2й и 5й документ с $y=3$ и $y=4$ соответственно.

Ответ: 6.5

Правильный ответ: 6.5

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

8. В задаче ранжирования важность корректного порядка неравномерна - гораздо важнее правильно упорядочить объекты в начале выдачи (наиболее релевантные по версии модели), чем в конце. Выберите меры качества, учитывающие этот эффект, где K =полное число объектов в выдаче:

☐ ☐ Precision@K

☐ ☐ доля дефектных (неверно упорядоченных) пар объектов при выдаче K объектов

☒ ☒ MeanAveragePrecision@K

☒ ☒ DCG@K

☒ ☒ AveragePrecision@K

☒ ☒ nDCG@K

Балл: 2.0

Комментарий к правильному ответу: