



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

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

Балл: 2.0

Комментарий к правильному ответу:

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

Балл: 2.0

Комментарий к правильному ответу:



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

Ответ: 0.45

Правильный ответ: 0.45

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

$$\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{4} + \frac{1}{5} = 0.25 + 0.2 = 0.45$$

4. 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 нормализованную меру $nDCG@7$, если релевантными оказались 2й и 5й документ с $y=3$ и $y=4$ соответственно. Ответ введите с точностью до 2 знаков после запятой

Ответ: 0.35

Правильный ответ: 0.35

Погрешность: 0.001

Балл: 2.0

Комментарий к правильному ответу:

5. 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:

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

6. 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



- ☐ ☐ даже небольшое сокращение ранга приводит к полной потере информации об индивидуальных особенностях пользователя с невозможностью в дальнейшем построить для него рекомендацию
- ☒ ☒ требуется заполнение пропущенных рейтингов в матрице рейтингов некоторыми числами, которые трактуются наравне с реально полученными рейтингами

Балл: 2.0

Комментарий к правильному ответу:

7. Почему на практике не используется

минимизация доли неверно упорядоченных объектов напрямую при обучении попарных (pairwise) методов градиентными методами оптимизации?

- ☐ ☐ для доли дефектных пар, в отличие от др. функций потерь, не очевиден выбор начального приближения параметров модели
- ☒ ☒ этот функционал является кусочно-постоянным, что не позволяет его минимизировать градиентными методами
- ☐ ☐ используемые на практике функции потерь вычисляются на порядок быстрее для пар объектов, составляющих минибатч

Балл: 2.0

Комментарий к правильному ответу:

8. Неявные оценки товаров (implicit feedback)

пользователями рекомендательной системы включают в себя:

- ☐ ☐ выставление лайка/дизлайка товару
- ☐ ☐ развернутый отзыв на товар
- ☒ ☒ факт добавления товара в корзину
- ☐ ☐ выставление оценки товару
- ☒ ☒ факт просмотра описания товара
- ☒ ☒ факт покупки товара

Балл: 2.0

Комментарий к правильному ответу: