



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
- ☒ ☒ существует неоднозначность выбора степени релевантности для соблюдения корректного порядка в ранжировании
- ☐ ☐ 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

**Балл: 2.0**

**Комментарий к правильному ответу:**

2. 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:**

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

3. Неявные оценки товаров (implicit feedback) пользователями рекомендательной системы включают в себя:

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



☐ ☐ выставление лайка/дизлайка товару

**Балл: 2.0**

**Комментарий к правильному ответу:**

4. 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:

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

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

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

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

**Балл: 2.0**

**Комментарий к правильному ответу:**

5. Рассмотрим меру Discounted Cumulative Gain (DCG) с функцией расчета выигрыша  $\text{Gain}(y) = 2^y - 1$  ( $y$ -степень релевантности) и функцией дисконтирования за более позднюю выдачу  $\text{Discount}(i) = 1/i$ , где  $i=1,2,3,\dots$  - номер позиции. Вычислите нормализованную меру  $\text{nDCG}@7$ , 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:**

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

**Ответ: 0.45**

**Правильный ответ: 0.45**

**Погрешность: 0.0**

**Балл: 2.0**

**Комментарий к правильному ответу:**

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



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

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

**Балл: 2.0**

**Комментарий к правильному ответу:**

8. 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: 0**

**Комментарий к правильному ответу:**