



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 модель с единственным скалярным признаком  $x$ , а  $y=+1$ , если  $x \geq A$  и  $y=-1$ , если  $x < A$ , где  $A$ -неизвестный параметр, функциональный вид зависимости известен,  $x$  и  $A$  принадлежат отрезку  $[0,1]$ . Пусть обучающую выборку мы можем формировать самостоятельно, назначая  $x$  и узнавая соответствующий  $y$ . Рассмотрим 3 метода построения обучающей выборки из  $N$  объектов: 1) равномерно делим  $[0,1]$  на  $N+1$  равных частей 2) выбираем  $x$   $N$  раз случайно и независимо из равномерного распределения на  $[0,1]$  3) объекты для разметки выбираются интеллектуально, чтобы обеспечить максимальную точность нахождения порога  $A$  за минимальное число последовательных разметок объектов (придумать как). Выберите верные погрешности нахождения порога  $A$  для каждой из 3х стратегий (т.е. 3 верных ответа):

☒ ☒ 1ая стратегия:  $O(1/N)$

☐ ☐ 1ая стратегия:  $O(1/N^2)$

☐ ☐ 2ая стратегия:  $O((\ln N)/N)$

☒ ☒ 3я стратегия:  $O(1/2^N)$

☐ ☐ 3я стратегия:  $O(1/N)$

☐ ☐ 2ая стратегия:  $O(1/2^N)$

☐ ☐ 3я стратегия:  $O(1/N^2)$

☐ ☐ 1ая стратегия:  $O((\ln N)/N)$

☐ ☐ 3я стратегия:  $O((\ln N)/N)$

☐ ☐ 1ая стратегия:  $O(1/2^N)$

☒ ☒ 2ая стратегия:  $O(1/N)$

☐ ☐ 2ая стратегия:  $O(1/N^2)$

**Балл: 2**

**Комментарий к правильному ответу:**

3я стратегия - деление отрезка с  $A$  каждый раз пополам. 2ая стратегия -

<https://math.stackexchange.com/questions/786392/expectation-of-minimum-of-n-i-i-d-uniform-random-variables>



2. Построение рекомендаций с помощью матричных разложений в общем случае обладает преимуществом, по сравнению с рекомендациями посредством обычного сокращенного сингулярного разложения (pure SVD) в том, что при настройке модели

- ☐ ☐ не учитываются пользователи и товары, у которых число рейтингов ниже заданного положительного порога
- ☐ ☒ появляется возможность аппроксимировать поведение пользователей матрицей более высокого ранга по сравнению с обычным сингулярным разложением
- ☒ ☐ используются только реально предоставленные рейтинги, а не все комбинации пользователь-товар в матрице рейтингов

**Балл: 0**

**Комментарий к правильному ответу:**

3. Выберите метод активного обучения, напрямую применимый к регрессии:

- ☐ ☐ entropy sampling
- ☒ ☒ expected model change
- ☐ ☐ least confident sampling
- ☐ ☐ margin sampling

**Балл: 2**

**Комментарий к правильному ответу:**

4. Рассмотрим две валидационных выборки: в первой объекты сэмпляются случайно из генеральной совокупности, а во второй - выбираются методом entropy sampling активного обучения. При этом модель обучается на обучающей выборке, полученной случайным сэмплением из генеральной совокупности и не дообучается на валидационных выборках - на них лишь тестируется качество модели. Как будут в общем случае связаны между собой точности модели на валидационных выборках, полученных каждым из способов?

- ☐ ☐ на второй выборке ожидается более высокая точность
- ☒ ☒ на второй выборке ожидается более низкая точность
- ☐ ☐ ожидается, что точности будут примерно соответствовать друг другу

Балл: 2

Комментарий к правильному ответу:

5. Факторизационные машины, в отличие от линейных моделей прогнозирования, позволяют учесть влияние на отклик

- ☒ ☒ попарных произведений различных признаков
- ☐ ☐ любых нелинейных преобразований от исходных признаков
- ☐ ☐ тригонометрических преобразований от исходных признаков ( $\sin(x)$ ,  $\cos(x)$ )
- ☐ ☐ квадратов исходных признаков

Балл: 2

Комментарий к правильному ответу:

6. Мера serendipity в рекомендательных системах показывает способность рекомендательной системы

- ☐ ☐ давать хорошие рекомендации для новых товаров, еще не рекомендовавшихся системой
- ☐ ☐ давать рекомендации, на которые пользователь будет часто кликать
- ☐ ☐ предоставлять разнообразные рекомендации, а не вариации одного и того же товара
- ☐ ☐ давать хорошие рекомендации для новых пользователей, еще не использовавших систему
- ☐ ☐ скрывать сомнительные рекомендации, которые пользователю могут не понравиться
- ☒ ☒ удивлять пользователя новыми нестандартными рекомендациями

Балл: 2

Комментарий к правильному ответу:

7. Проблема холодного старта (cold start) в рекомендательных системах заключается

- ☐ ☐ в устойчивом смещении оценок пользователей в зависимости от их личной шкалы оценивания
- ☒ ☒ в сложности построения рекомендаций новым пользователям и товарам, о рейтингах которых еще не собрано подробной статистики
- ☐ ☐ в смещении оценок товаров, вызванных фальшивыми оценками производителей этих товаров либо их конкурентами
- ☐ ☐ в сложности оценивания перепараметризованной рекомендательной системы

Балл: 2

Комментарий к правильному ответу:



8. Может ли модель, обученная методом активного обучения на  $N$  объектах, дать точность на независимой валидационной выборке, репрезентативной для генеральной совокупности, хуже, чем модель, обученная на  $N$  случайно выбранных объектах?

☒ ☒ да

☐ ☐ нет

**Балл: 2**

**Комментарий к правильному ответу:**