



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 две валидационных выборки: в первой объекты сэмплируются случайно из генеральной совокупности, а во второй - выбираются методом entropy sampling активного обучения. При этом модель обучается на обучающей выборке, полученной случайным сэмплированием 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?

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

Балл: 2

Комментарий к правильному ответу:

2. Пусть мы используем ансамбль моделей для прогнозирования в задаче регрессии. Предложите корректный вариант активного обучения в этом случае (вариант vote entropy - но для регрессии). Нужно выбирать объекты для доразметки, на которых

- ☐ ☐ среднее прогнозов базовыми моделями ансамбля ниже
- ☐ ☐ среднее прогнозов базовыми моделями ансамбля выше
- ☐ ☐ дисперсия прогнозов базовых моделей ансамбля ниже
- ☒ ☒ дисперсия прогнозов базовых моделей ансамбля выше

Балл: 2

Комментарий к правильному ответу:

3. 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 выборки 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):

- ☐ ☐ 2ая стратегия: $O(1/N^2)$
- ☐ ☐ 3я стратегия: $O(1/N^2)$
- ☐ ☐ 3я стратегия: $O(1/N)$
- ☐ ☐ 3я стратегия: $O((\ln N)/N)$
- ☒ ☒ 1ая стратегия: $O(1/N)$
- ☒ ☒ 2ая стратегия: $O(1/N)$
- ☐ ☐ 2ая стратегия: $O(1/2^N)$
- ☐ ☐ 1ая стратегия: $O((\ln N)/N)$
- ☐ ☐ 1ая стратегия: $O(1/N^2)$
- ☐ ☐ 1ая стратегия: $O(1/2^N)$
- ☒ ☒ 3я стратегия: $O(1/2^N)$
- ☐ ☐ 2ая стратегия: $O((\ln N)/N)$

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>

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



☒ ☒ expected model change

☐ ☐ margin sampling

☐ ☐ least confident sampling

☐ ☐ entropy sampling

Балл: 2

Комментарий к правильному ответу:

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

☐ ☐ тригонометрических преобразований от исходных признаков ($\sin(x)$, $\cos(x)$)

☒ ☒ попарных произведений различных признаков

☐ ☐ любых нелинейных преобразований от исходных признаков

☐ ☐ квадратов исходных признаков

Балл: 2

Комментарий к правильному ответу:

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

☐ ☐ в устойчивом смещении оценок пользователей в зависимости от их личной шкалы оценивания

☐ ☐ в смещении оценок товаров, вызванных фальшивыми оценками производителей этих товаров либо их конкурентами

☐ ☐ в сложности оценивания перепараметризованной рекомендательной системы

☒ ☒ в сложности построения рекомендаций новым пользователям и товарам, о рейтингах которых еще не собрано подробной статистики

Балл: 2

Комментарий к правильному ответу:

7. Допустим, пользователю в интернет-магазине нужно быстро перестраивать рекомендации в зависимости от текущего содержания его корзины, которую он быстро наполняет. Какой алгоритм вычислительно проще для этой задачи



☒ ☒ item-based

☐ ☐ нет разницы в производительности этих методов

☐ ☐ user-based

Балл: 2

Комментарий к правильному ответу:

При изменении профиля пользователя user-user требует пересчета похожести пользователя на всех остальных, а профиль товара не сильно меняется.

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: