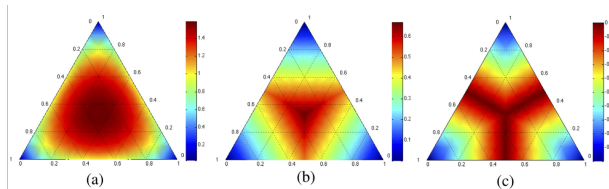


В тестовых заданиях первая галочка — правильный ответ, вторая галочка — выбранный ответ. Цвет обозначает, правильно ли в данном пункте поставлена галочка. Если все пункты верные (галочки совпадают / все пункты зеленые), то за задание ставится полный балл, в противном случае ставится 0 баллов.

1. Пусть имеется задача классификации на 3 класса. Рассмотрим тепловые карты степени приоритетности объектов для разметки методами least confident sampling, margin sampling, entropy sampling. Какая из карт соответствует какому методу? Левый нижний угол - объект с вероятностями классов согласно модели  $(1,0,0)$ , верхний -  $(0,1,0)$ , правый нижний -  $(0,0,1)$ . Остальные точки соответствуют промежуточным вероятностям классов  $(p(1), p(2), p(3))$  линейной интерполяцией.



- ☐ ☐ (a) - margin sampling, (b) - least confident sampling, (c) - entropy sampling
- ☐ ☐ (a) - least confident sampling, (b) - margin sampling, (c) - entropy sampling
- ☐ ☐ (a) - least confident sampling, (b) - entropy sampling, (c) - margin sampling
- ☒ ☒ (a) - entropy sampling, (b) - least confident sampling, (c) - margin sampling
- ☐ ☐ (a) - entropy sampling, (b) - margin sampling, (c) - least confident sampling
- ☐ ☐ (a) - margin sampling, (b) - entropy sampling, (c) - least confident sampling

**Балл: 2**

**Комментарий к правильному ответу:**

У margin  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  так же приоритетен, как  $(\frac{1}{2}, \frac{1}{2}, 0)$ ,  $(0, \frac{1}{2}, \frac{1}{2})$  и т.д. У least confident - линии уровня  $p(i)=\text{const}$  - отрезки напротив  $i$ -й вершины.

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

- ☐ ☐ в сложности оценивания перепараметризованной рекомендательной системы
- ☐ ☐ 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 построения рекомендаций новым пользователям и товарам, о рейтингах которых еще не собрано подробной статистики

**Балл: 2**

**Комментарий к правильному ответу:**

3. 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

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

**Балл: 0**

**Комментарий к правильному ответу:**

4. Задача активного обучения состоит в том, чтобы

- ☐ ☐ активнее использовать существующие размеченные объекты, неоднократно возвращаясь к ним в процессе обучения
- ☒ ☒ интеллектуально выбирать объекты из неразмеченной части выборки (метки неизвестны) для их доразметки, потом дообучать модель на расширенной обучающей выборке (с известными метками)
- ☐ ☐ использовать неразмеченную часть выборки (отклики неизвестны) для донастройки модели без узнавания их реальных откликов

**Балл: 2.0**

**Комментарий к правильному ответу:**

5. 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

- ☐ ☐ user-based



☐ ☐ нет разницы в производительности этих методов

☒ ☒ item-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.

6. 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

☒ ☒ дисперсия прогнозов базовых моделей ансамбля выше

☐ ☐ среднее прогнозов базовыми моделями ансамбля выше

☐ ☐ дисперсия прогнозов базовых моделей ансамбля ниже

**Балл: 2**

**Комментарий к правильному ответу:**

7. 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):

☒ ☒ 2ая стратегия:  $O(1/N)$



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

**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>

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

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

**Балл: 2**

**Комментарий к правильному ответу:**