



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 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

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

Балл: 2

Комментарий к правильному ответу:

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

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

Балл: 2

Комментарий к правильному ответу:

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

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

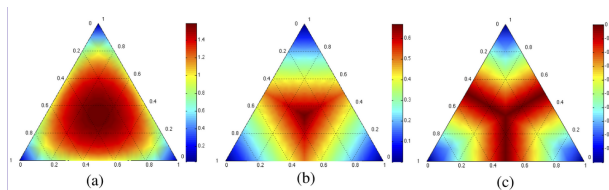
Балл: 2**Комментарий к правильному ответу:**

4. Задача активного обучения состоит в том, чтобы

- ☐ ☐ использовать неразмеченную часть выборки (отклики неизвестны) для донастройки модели без узнавания их реальных откликов
- ☐ ☐ 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)

Балл: 2.0**Комментарий к правильному ответу:**

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

6. 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.

7. 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

- ☒ ☐ используются только реально проставленные рейтинги, а не все комбинации пользователь-товар в матрице рейтингов
- ☐ ☒ 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

Балл: 0

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: