

Анна



## Комментарии к тесту

Машинное обучение 2022-2023

Материалы (/course/4/info)

- 
1. Введение в Python.1-5 (/course/4/task/1)
  2. Введение в Python.6 (/course/4/task/2)
  3. Введение в Python.7 (/course/4/task/3)
  4. numpy-pandas-matplotlib (unit-tests) (/course/4/task/4)
  5. Тест №1 (/course/4/task/5)
  6. numpy-pandas-matplotlib (notebook) (/course/4/task/6)
  7. KNN (unit-tests) (/course/4/task/7)
  8. KNN (notebook) (/course/4/task/8)
  9. Тест №2 (/course/4/task/9)
  10. Linear models (unit-tests) (/course/4/task/10)
  11. Linear models (notebook) (/course/4/task/11)
  12. Тест №3 (/course/4/task/12)
  13. Основы SVM (ML) (/course/4/task/13)
  14. Основы SVM (notebook) (/course/4/task/14)

В множественном выборе есть хотя бы один правильный ответ. Их может быть как один, так и несколько. Полный балл будет ставиться только в том случае, если выбраны все верные ответы и не выбраны все неверные ответы.

В заданиях, где нужно ввести число, в данном тесте, ответ всегда целочисленный. Засчитываться будет только точное совпадение

- При подсчете разбиений на тестовую и обучающую выборки порядок объектов не важен.
- Метод ближайших центроидов осуществляет классификацию, в которой усреднением по объектам класса вычисляется центроид каждого класса, а для нового объекта назначается тот класс, центроид которого ближе.
- В метрических методах используется Евклидова ф-ция расстояния, если не оговорено иначе.
- Обозначим  $N$  - число объектов обучающей выборки,  $D$ -число признаков,  $C$ -число классов.
- Под “простотой” и “сложностью” моделей понимается не сложность алгоритма или вычислений, а ее негибкость и гибкость соответственно, т.е. способность адаптироваться к данным обучающей выборки. Если речь идет о сложности вычислений, то это называется “вычислительной простотой” и “вычислительной сложностью”.
- Критерий качества - функция, максимизируя которую выбирается наилучшая модель. Функция потерь - функция, минимизируя которую выбирается наилучшая модель.
- Отличие функции близости и функции расстояния на примере метода  $K$  ближайших соседей: ближайшие соседи сортируются по возрастанию ф-ции расстояния либо сортируются по убыванию ф-ции близости.
- $x^{(i)}, z^{(i)}$  --  $i$ -й признак объекта  $x(z)$ .

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.

## Анна

Машинное обучение 2022-2023

Материалы (/course/4/info)

---

1. Введение в Python.1-5 (/course/4/task/1)
2. Введение в Python.6 (/course/4/task/2)
3. Введение в Python.7 (/course/4/task/3)
4. numpy-pandas-matplotlib (unit-tests) (/course/4/task/4)
5. Тест №1 (/course/4/task/5)
6. numpy-pandas-matplotlib (notebook) (/course/4/task/6)
7. KNN (unit-tests) (/course/4/task/7)
8. KNN (notebook) (/course/4/task/8)
9. Тест №2 (/course/4/task/9)
10. Linear models (unit-tests) (/course/4/task/10)
11. Linear models (notebook) (/course/4/task/11)
12. Тест №3 (/course/4/task/12)
13. Основы SVM (ML) (/course/4/task/13)
14. Основы SVM (notebook) (/course/4/task/14)

1. Выберите функции расстояния, которые сделают прогнозы методом  $K$  ближайших соседей независимыми к масштабированию признаков:

☐ ☐ Евклидово  $\sqrt{\sum_{i=1}^D (x^{(i)} - z^{(i)})^2}$

☒ ☒ Канберра  $\frac{1}{D} \sum_{i=1}^D \frac{|x^{(i)} - z^{(i)}|}{x^{(i)} + z^{(i)}}$

☐ ☐  $L_p = \sqrt[p]{\sum_{i=1}^D (x^{(i)} - z^{(i)})^p}$

☐ ☐ Ланса-Уильямса  $\frac{\sum_{i=1}^D |x^{(i)} - z^{(i)}|}{\sum_{i=1}^D x^{(i)} + z^{(i)}}$

**Балл: 2.0**

**Комментарий к правильному ответу:**

2. В машинном обучении решается задача предсказать...

☐ ☐ ничего из перечисленного

☒ ☒ по признакам отклики (целевую переменную)

☐ ☐ по откликам признаки

☐ ☐ признаки и отклики (целевые переменные)

**Балл: 2.0**

**Комментарий к правильному ответу:**

3. Будем судить о похожести/непохожести объектов по косинусу угла между их векторами признаков представлений (косинусная мера). Эта мера будет показывать...

☒ ☒ степень близости между объектами (больше-ближе)

☐ ☐ степень расстояния (непохожести) между объектами (больше-дальше)

Анна

Машинное обучение 2022-2023

Материалы (/course/4/info)

- 
1. Введение в Python.1-5 (/course/4/task/1)
  2. Введение в Python.6 (/course/4/task/2)
  3. Введение в Python.7 (/course/4/task/3)
  4. numpy-pandas-matplotlib (unit-tests) (/course/4/task/4)
  5. Тест №1 (/course/4/task/5)
  6. numpy-pandas-matplotlib (notebook) (/course/4/task/6)
  7. KNN (unit-tests) (/course/4/task/7)
  8. KNN (notebook) (/course/4/task/8)
  9. Тест №2 (/course/4/task/9)
  10. Linear models (unit-tests) (/course/4/task/10)
  11. Linear models (notebook) (/course/4/task/11)
  12. Тест №3 (/course/4/task/12)
  13. Основы SVM (ML) (/course/4/task/13)
  14. Основы SVM (notebook) (/course/4/task/14)

Балл: 2.0

Комментарий к правильному ответу:

4. Выбор параметров для методов машинного обучения достигается минимизацией...

☐ ☐ теоретического риска☒ ☒ эмпирического риска

Балл: 2.0

Комментарий к правильному ответу:

5. В задаче классификации нуклеотидных последовательностей объекты задаются строками из четырёх букв А, G, T и C. В качестве числовых признаков объектов решено использовать частоты букв, частоты всевозможных биграмм (AA, AG, ..., CT, CC) и частоты всевозможных триграмм (AAA, AAG, ..., CCT, CCC). Сколько всего признаков?

Ответ: 84

Правильный ответ: 84.0

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

 $4 + 4 \cdot 4 + 4 \cdot 4 \cdot 4$ 

6. При взвешенном обобщении метода  $K$  ближайших соседей более близким объектам нужно сопоставлять...

☐ ☐ более низкий вес☒ ☒ более высокий вес

Балл: 2.0

## Анна

Машинное обучение 2022-2023

Материалы (/course/4/info)

- 
1. Введение в Python.1-5 (/course/4/task/1)
  2. Введение в Python.6 (/course/4/task/2)
  3. Введение в Python.7 (/course/4/task/3)
  4. numpy-pandas-matplotlib (unit-tests) (/course/4/task/4)
  5. Тест №1 (/course/4/task/5)
  6. numpy-pandas-matplotlib (notebook) (/course/4/task/6)
  7. KNN (unit-tests) (/course/4/task/7)
  8. KNN (notebook) (/course/4/task/8)
  9. Тест №2 (/course/4/task/9)
  10. Linear models (unit-tests) (/course/4/task/10)
  11. Linear models (notebook) (/course/4/task/11)
  12. Тест №3 (/course/4/task/12)
  13. Основы SVM (ML) (/course/4/task/13)
  14. Основы SVM (notebook) (/course/4/task/14)

## Комментарий к правильному ответу:

7. Дана выборка из 10 объектов. Сколькими способами её можно разбить на две непустые выборки, обучающую и тестовую? (порядок следования объектов не важен)

**Ответ:** 1022**Правильный ответ:** 1022.0**Погрешность:** 0.0**Балл:** 2.0**Комментарий к правильному ответу:**

$2^{10} - 2$ . Это количество всевозможных последовательностей из 0 и 1 длины 10 (например, 0 - принадлежность объекта обучающей выборке, 1 - тестовой). И вычитаем две последовательности (из всех 0 и всех 1), так как мы хотим учитывать только непустые разбиения

8. Нам нужно получить оценку эмпирического риска на тестовой выборке. Сравните подход кросс-валидации и отдельной валидационной выборки, полученной как фрагмент обучающей. Какой метод ближе оценит качество итоговой модели на тестовой выборке, обученной по всей обучающей выборке?

☐ ☐ Оценка по валидационной выборке☒ ☒ Оценка кросс-валидации**Балл:** 2.0**Комментарий к правильному ответу:**