



## Комментарии к тесту

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 функция 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 функции расстояния либо сортируются по убыванию функции 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.

1. Если в качестве оценки ожидаемого эмпирического риска на тестовой выборке использовать эмпирический риск на обучающей выборке, то в большинстве случаев получим...

- ☐ ☐ реальную оценку
- ☐ ☐ пессимистическую оценку (хуже, чем на самом деле)
- ☒ ☒ оптимистическую оценку (лучше, чем на самом деле)

**Балл:** 2.0

**Комментарий к правильному ответу:**

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)
15. Тест №4 (/course/4/task/15)

Таблица результатов (/course/4/standings)

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)
15. Тест №4 (/course/4/task/15)

2. 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)

**Балл:** 2.0

**Комментарий к правильному ответу:**

3. При взвешенном обобщении метода  $K$  ближайших соседей с весами равными обратным величинам расстояний до объектов, прогноз в некоторой достаточно малой окрестности объектов-выбросов (имеющих класс не совпадающий с классом всех окружающих объектов):

- ☒ ☒ будет совпадать с классом выброса
- ☐ ☐ будет совпадать с классом окружающих объектов

**Балл:** 2.0

**Комментарий к правильному ответу:**

4. Рассмотрим многоклассовую классификацию методом  $K$  ближайших соседей с равномерными весами. Может ли возникать ситуация, что два класса набирают одинаковый рейтинг (равные дискриминантные функции)?

- ☐ ☐ Не может для нечетного  $K$
- ☒ ☒ Может даже для нечетного  $K$
- ☐ ☐ Не может

**Балл:** 2.0

**Комментарий к правильному ответу:**

5. Дана выборка из 10 объектов. Сколькими способами её можно разбить на две непустые

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)15. Тест №4  
(/course/4/task/15)Таблица результатов  
(/course/4/standings)

выборки, обучающую и тестовую? (порядок следования объектов не важен)

**Ответ:** 1765**Правильный ответ:** 1022.0**Погрешность:** 0.0**Балл:** 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

6. Дана выборка из 10 объектов. Сколькими способами её можно разбить на две выборки, обучающую и тестовую, так, чтобы в тестовой оказалось ровно два объекта?

**Ответ:** 45**Правильный ответ:** 45.0**Погрешность:** 0.0**Балл:** 2.0**Комментарий к правильному ответу:**

$C_{10}^2$  --- количество способов выбрать два объекта из 10, которых мы включим в тестовую выборку.

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

☐ ☐ Евклидово  $\sqrt{\sum_{i=1}^D (x^{(i)} - z^{(i)})^2}$ 
☐ ☐  $L_p = \sqrt[p]{\sum_{i=1}^D (x^{(i)} - z^{(i)})^p}$ 
☒ ☒ Канберра  $\frac{1}{D} \sum_{i=1}^D \frac{|x^{(i)} - z^{(i)}|}{|x^{(i)} + z^{(i)}|}$ 
☐ ☒ Ланса-Уильямса  $\frac{\sum_{i=1}^D |x^{(i)} - z^{(i)}|}{\sum_{i=1}^D |x^{(i)} + z^{(i)}|}$ 
**Балл:** 0**Комментарий к правильному ответу:**

8. Нам нужно получить оценку эмпирического риска на тестовой выборке. Сравните подход кросс-валидации и



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)

15. Тест №4  
(/course/4/task/15)

Таблица результатов  
(/course/4/standings)

отдельной валидационной выборки, полученной как фрагмент обучающей. Какой метод ближе оценит качество итоговой модели на тестовой выборке, обученной по всей обучающей выборке?

☒ ☒ Оценка кросс-валидации

☐ ☐ Оценка по валидационной выборке

**Балл:** 2.0

**Комментарий к правильному ответу:**