



## 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 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.

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

☒ ☒ более высокий вес

☐ ☐ более низкий вес

**Балл: 2.0**

**Комментарий к правильному ответу:**

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. 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:**

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

☒ ☒ Может даже для нечетного  $K$

☐ ☐ Не может

☐ ☐ Не может для нечетного  $K$

**Балл: 2.0**

**Комментарий к правильному ответу:**

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

☒ ☒ эмпирического риска

☐ ☐ теоретического риска

**Балл: 2.0**

**Комментарий к правильному ответу:**

6. 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

7. В машинном обучении решается задача предсказать...

- ☐ ☐ по откликам признаки
- ☒ ☒ по признакам отклики (целевую переменную)
- ☐ ☒ признаки и отклики (целевые переменные)
- ☐ ☐ ничего из перечисленного

**Балл:** 0

**Комментарий к правильному ответу:**

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

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

**Балл:** 2.0

**Комментарий к правильному ответу:**