



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. 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 кросс-валидации

☐ ☐ Оценка по валидационной выборке

Балл: 2.0
Комментарий к правильному ответу:

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

☒ ☒ Канберра $\frac{1}{D} \sum_{i=1}^D \frac{|x^{(i)} - z^{(i)}|}{|x^{(i)} + z^{(i)}|}$
☐ ☐ Евклидово $\sqrt{\sum_{i=1}^D (x^{(i)} - z^{(i)})^2}$
☐ ☐ Ланса-Уильямса $\frac{\sum_{i=1}^D |x^{(i)} - z^{(i)}|}{\sum_{i=1}^D |x^{(i)} + z^{(i)}|}$
☐ ☐ $L_p = \sqrt[p]{\sum_{i=1}^D (x^{(i)} - z^{(i)})^p}$
Балл: 2.0
Комментарий к правильному ответу:

4. 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.

5. В задаче классификации с двумя классами дана выборка из 10 объектов, по 5 объектов в каждом классе. Сколькими способами её можно разбить на две выборки, обучающую и тестовую, так, чтобы в тестовой оказалось ровно два объекта и они принадлежали разным классам?

Ответ: 25
Правильный ответ: 25.0
Погрешность: 0.0
Балл: 2.0
Комментарий к правильному ответу:

(5 способов выбрать 1 объект класса А) · (5 способов выбрать 1 объект класса Б)



6. В задаче классификации с двумя классами дана выборка из 8 объектов, по 4 объекта в каждом классе. Сколькими способами её можно разбить на обучающую и тестовую так, чтобы в обеих оказалось по 2 объекта каждого класса?

Ответ: 36

Правильный ответ: 36.0

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

$(C_4^2 \text{ способов выбрать 2 объекта класса А}) \cdot (C_4^2 \text{ способов выбрать 2 объекта класса Б}) = 6 \cdot 6$

7. 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 окружающих объектов

☒ ☒ будет совпадать с классом выброса

Балл: 2.0

Комментарий к правильному ответу:

8. 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

☐ ☒ Не может для нечетного K

☒ ☐ Может даже для нечетного K

Балл: 0

Комментарий к правильному ответу: