



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. Прогнозы методом K ближайших соседей:

- ☒ ☒ зависят от масштабирования признаков
- ☐ ☐ не зависят от масштабирования признаков

Балл: 2.0

Комментарий к правильному ответу:

2. 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:

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

3. 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

4. В задаче классификации нуклеотидных последовательностей объекты задаются строками из четырёх букв А, G, Т и С. В качестве числовых признаков объектов решено использовать частоты букв, частоты всевозможных биграмм (АА, AG, ..., СТ, СС) и частоты всевозможных триграмм (ААА, ААG, ..., ССТ, ССС). Сколько всего признаков?

Ответ: 84

Правильный ответ: 84.0

Погрешность: 0.0

Балл: 2.0

Комментарий к правильному ответу:

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

5. 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:



6. 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
- ☐ ☐ Не может
- ☒ ☒ Может даже для нечетного K

Балл: 2.0

Комментарий к правильному ответу:

7. 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:

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

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

Балл: 0

Комментарий к правильному ответу: