

Лабораторная работа №4 - Классификация

Набор данных: lab4_5.csv

Для построения ROC-кривой и расчета AUC в случае множества классов, можно построить для одного класса по схеме OvA (для получения вероятностей использовать метод классификатора `predict_proba`).

Для расчета Precision, Recall, F1 счета можно использовать

[sklearn.metrics.precision_recall_fscore_support](#) . Для нескольких классов усреднение можете выбрать самостоятельно.

1. Загрузка данных

- 1.1. Загрузите данные вашего варианта. Учтите, что метки классов являются текстом.
- 1.2. Визуализируйте данные при помощи диаграммы рассеяния с выделением различных классов ней.
- 1.3. Оцените сбалансированность классов
- 1.4. Проведите предобработку данных при необходимости
- 1.5. Разделите выборку на обучающую и тестовую. На обучающей проводите обучение модели, на тестовой расчет значений метрик.

2. kNN

- 2.1. Проведите классификацию методом k ближайших соседей, подобрав параметры: количество соседей, необходимость взвешивани. Постройте графики зависимости точности (Accuracy) в зависимости от количества соседей (по 1-й линии для взвешенного и не взвешенного метода)
- 2.2. Рассчитайте значения Precision, Recall, F1 для полученных результатов. Сопоставьте с таблицей ошибок.

3. Логистическая регрессия

- 3.1. Проведите классификацию методом логистической регрессии при различных параметрах: без регуляризации, с l1 регуляризацией, с l2 регуляризацией. Постройте столбчатую диаграмму зависимости точности (Accuracy) от наличия регуляризации. Дальнейшие пункты 3.* выполняйте для лучшего параметра.
- 3.2. Рассчитайте значения Precision, Recall, F1 для полученных результатов. Сопоставьте с таблицей ошибок.

4. Метод опорных векторов

- 4.1. Проведите классификацию методом опорных векторах при различных параметрах ядра: "linear", "poly" (нужно выбрать степень), "rbf". Постройте столбчатую диаграмму зависимости точности (Accuracy) от вида ядра. Дальнейшие пункты 4.* выполняйте для лучшего параметра.

- 4.2. Рассчитайте значения Precision, Recall, F1 для полученных результатов. Сопоставьте с таблицей ошибок.

5. Решающие деревья

- 5.1. Проведите классификацию используя решающие деревья, подобрав параметры при которых получается лучшее обобщение (максимальная глубина/максимальное количество листьев/метрика загрязнения и т.д.).
- 5.2. Рассчитайте значения Precision, Recall, F1 для полученных результатов. Сопоставьте с таблицей ошибок.

6. Выбор классификатора

- 6.1. Постройте таблицу с метриками Precision, Recall, AUC для полученных результатов каждым классификатором.
- 6.2. Сделайте выводы о том, какие классификаторы лучше всего подходят для вашего набора данных и в каких случаях.