

Многоклассовая классификация. Подбор параметров по сетке

Практическое задание для самостоятельного выполнения

Задание.

1. Используя инструментарий модуля *datasets*, импортировать встроенный набор данных *iris* (часто используется авторами учебников и учебных курсов по машинному обучению в качестве учебного набора данных; содержит информацию о 150 объектах – цветах ириса, которые на основании значений четырех признаков необходимо отнести к одному из трех сортов).
Формат команды: `<имя набора данных> = datasets.load_iris()`
2. В результате выполнения п. 1 должен быть создан объект, по структуре подобный словарю. Чтобы убедиться в этом и получить представление о структуре набора данных, вывести наименования ключей набора.
3. Вывести значение ключа `DESCR` (содержит описание набора данных) и изучить описание набора.
4. Изучить значения остальных ключей и определить ключи, в которых хранятся значения признаков и метки классов. Далее вся работа – только со значениями этих ключей.
5. Выполнить разбиение набора данных (значения признаков и метки классов) на обучающую и тестовую выборки в соотношении 70/30.
6. Создать модель линейной классификации *SGDClassifier* и вывести список доступных параметров модели.
7. Создать сетку параметров, включающую 4 вида функции потерь, два типа регуляризаторов, 5 значений коэффициента регуляризации от 0.0001 до 0.001, и число итераций от 5 до 10 с шагом 1.
8. Создать объект *GridSearchCV*, передать ему созданный ранее классификатор и сетку параметров и обучить его. В качестве метрики использовать *accuracy*. Предусмотреть вывод времени, затраченного на перебор по сетке (команда `%%time`).
9. Вывести лучший классификатор, лучший набор параметров и оценку лучшего классификатора в соответствии с заданной метрикой. Получить оценки алгоритма по последним 10 наборам параметров.
10. Проанализировать полученные результаты, сделать выводы.
11. Организовать случайный поиск по сетке. Вывести время, затраченное на случайный перебор.
12. Вывести те же показатели, что и в п. 9. Сопоставить полученные результаты и время на перебор с результатами, полученными при выполнении п. 9.
13. Сделать выводы, сформулировать рекомендации.