

## РАБОТА С МЕТОДОМ ОПОЛНЫХ ВЕРКОРОВ (SVM)

Задание основано на применении пакета Python Scikit-learn для изучения метода опорных векторов.

Цели практики:

- Научиться работать с методом опорных векторов (SVM)
- Овладеть навыками нахождения наиболее важных объектов выборки

Начало работы

Метод опорных векторов реализован в классе **sklearn.svm.SVC**.

Основными параметрами этого класса являются коэффициент **C** и тип ядра **kernel**.

В данной задаче мы будем использовать линейное ядро - для этого нужно задать значение параметра **kernel='linear'**.

Индексы опорных объектов обученного классификатора хранятся в поле **support\_**

### Задание

1. Загрузите выборку из файла `svm-data.csv`. В нем записана двумерная выборка (целевая переменная указана в первом столбце, признаки - во втором и третьем).
2. Обучите классификатор с линейным ядром, параметром  $C = 100000$  и `random_state=241`.

Такое значение параметра нужно использовать, чтобы убедиться, что SVM работает с выборкой как с линейно разделимой. При более низких значениях параметра алгоритм будет настраиваться с учетом слагаемого в функционале, штрафующего за маленькие отступы, из-за чего результат может не совпасть с решением классической задачи SVM для линейно разделимой выборки.

3. Найдите номера объектов, которые являются опорными (нумерация с единицы). Они будут являться ответом на задание.

Обратите внимание, что в качестве ответа нужно привести номера объектов в возрастающем порядке через запятую. *Нумерация начинается с 1.*

**Ответ на каждое задание** — текстовый файл, содержащий ответ в первой строчке (`#.txt`). Обратите внимание, что отправляемые файлы не должны содержать перевод строки в конце.

## ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

Метод опорных векторов (**Support Vector Machine, SVM**) – один из видов линейных классификаторов. Функционал, который он оптимизирует, направлен на максимизацию ширины разделяющей полосы между классами.

Из теории статистического обучения известно, что эта ширина тесно связана с обобщающей способностью алгоритма, а ее максимизация позволяет бороться с переобучением.

Метод опорных векторов имеет еще одну особенность. Если преобразовать его оптимизационную задачу, то окажется, что итоговый классификатор можно представить как взвешенную сумму скалярных произведений данного объекта на объекты обучающей выборки:

По сути, алгоритм делает предсказания на основе сходства нового объекта с объектами обучающей выборки. При этом, как правило, далеко не все коэффициенты оказываются ненулевыми.

Это означает, что классификация делается на основе сходства лишь с частью обучающих объектов. *Такие объекты называются опорными.*