

## Лабораторная работа №4

### «Метод опорных векторов (Support Vector Machine, SVM)»

**Цель работы:** получить практические навыки работы с методом опорных векторов; познакомиться с возможностями интерактивной оболочки Jupyter Notebook и библиотеки scikit-learn.

#### Теоретический материал

Метод опорных векторов (Support Vector Machine) – алгоритм машинного обучения, который может быть использован, как для решения задач классификации, так и регрессии. В этом алгоритме каждый объект представляет собой точку в  $n$ -мерном пространстве (где  $n$ -число атрибутов объекта). Алгоритм относится к методу обучения с учителем, т.к. изначально имеется обучающая выборка, в которой заранее известна принадлежность объектов к классам.

Идея SVM заключается в следующем. Пусть имеется множество точек, разбитое на два класса (рис. 1). Проведем линию, разделяющую эти классы. В дальнейшем все новые точки будут классифицированы так: если точка находится выше разделяющей прямой, то она принадлежит к классу А, если ниже – к классу В.

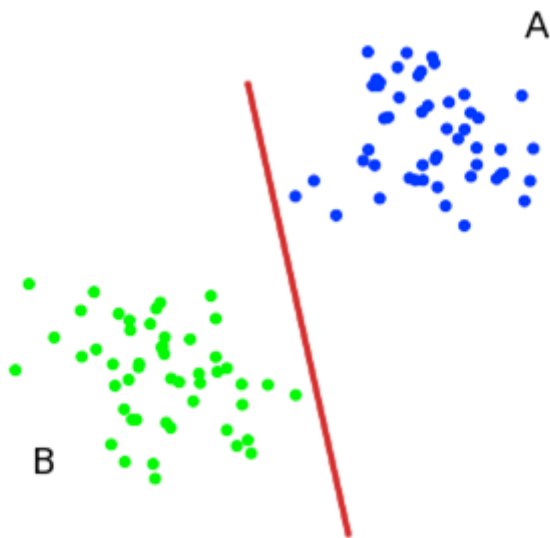


Рисунок 1 – Множество, разбитое на два класса

Однако часто возникают ситуации, когда линейно разделить классы невозможно (рис. 2). Для решения этой проблемы, все объекты из обучающей выборки с помощью специальных функций отображения вкладываются в пространство большей размерности, в котором элементы уже будут линейно разделимы.

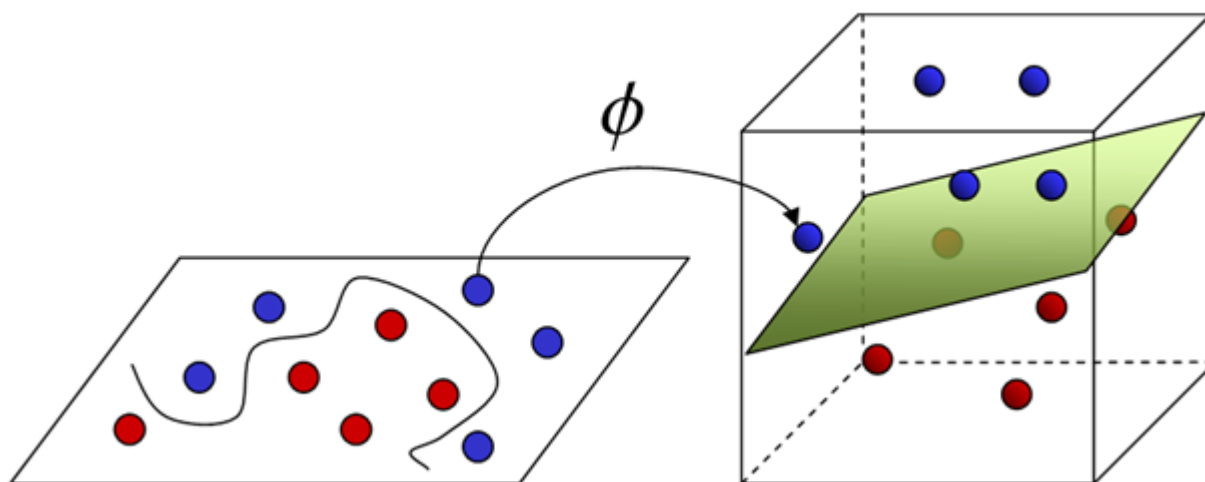


Рисунок 2 – Перевод значений обучающей выборки в пространство большей размерности

Т.к. объекты зачастую содержат множество признаков, то для разделения классов строится не линия, а гиперплоскость. Классификатор не обязательно должен быть линейным. С помощью функций ядра можно создать нелинейный классификатор для нахождения плоскостей максимальной разности. Наиболее распространенные ядра: полиномиальное, радиальная базисная функция, радиальная базисная функция Гаусса и сигмоида.

### Ход работы

1. Прочитать теоретическую часть по методу опорных векторов (Дополнительно: [https://github.com/artemvirused/ITMO\\_ML\\_Labs/blob/master/Lab4/ML\\_lab4\\_theory.ipynb](https://github.com/artemvirused/ITMO_ML_Labs/blob/master/Lab4/ML_lab4_theory.ipynb)).
2. Описать структуру исходных данных для своего набора:
  - а. общие характеристики массива данных: предметная область, количество записей
  - б. входные параметры: названия и типы
  - с. выходной класс: название и значения
3. Осуществить ряд экспериментов по классификации, используя SVM с различными параметрами (функции ядра и пр.), и занести результаты в сравнительную таблицу.
4. Выбрать оптимальные параметры и сформировать вывод о применимости метода опорных векторов к задаче классификации для своего набора данных.
5. Отчет оформить в виде блокнота Jupyter Notebook (File → Download as → HTML (.html))