

## Лабораторная работа №5.

Рассматривается задача бинарной классификации. Пусть заданы два класса, распределение которых является многомерным нормальным ( $\mu^+ \in \mathbb{R}^2$  – параметр положительного класса (+1),  $\mu^-$  – отрицательного (-1),  $\Sigma \in \mathbb{R}^{2 \times 2}$  – общая для классов ковариационная матрица).

1. Подобрать параметры распределений таким образом, чтобы классы немного пересекались (нельзя было разделить их линейной функцией).
2. Сгенерировать по  $N = 30$  объектов каждого класса. Разделить данные каждого класса на 3 равные по объему выборки (тренировочную, валидационную и тестовую).
3. Реализовать и обучить модель логистической регрессии (логику написать самим). Построить на графике полученную разделяющую поверхность. Для нескольких объектов из обоих классов вывести оценки вероятностей их распределения по классам.
4. Определить оптимальный порог  $t$ , при котором на валидационной выборке определяются не менее 60% от поступающих положительных объектов:

$$\hat{p}(y = +1|x) \equiv a(x) > t.$$

5. На тестовой выборке оценить качество полученного алгоритма, рассчитав метрики (расчет метрик реализовать самим):
  - 5.1. Accuracy;
  - 5.2. Матрица ошибок;
  - 5.3. Precision, Recall;
  - 5.4. ROC – кривая, ROC-AUC.