**Федеральное агентство по образованию**Государственное образовательное учреждение высшего профессионального   
образования **«Тихоокеанский Государственный университет»**

Факультет компьютерных и фундаментальных наук

Кафедра ПОВТАС

**Лабораторная работа №1**

по дисциплине: «Методы машинного обучения»

на тему: «Расчет коэффициентов разделяющей линии и вычисление отступа (margin) для объектов разных классов»  
Вариант №4

Выполнил: студент группы ПИИ(м)-21

Забавин А.С.

Проверил: преподаватель кафедры ПОВТАС

Тормозов В.С.

# Постановка задачи

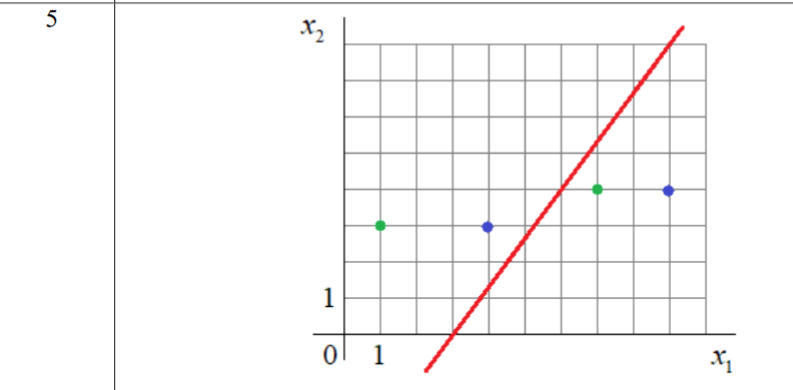
**Цель работы**: научиться вычислять коэффициенты разделяющей линии и  
величину отступа (margin) при бинарной классификации объектов.

**Задания на лабораторную работу (5 вариант)**

1. Используя рисунок своего варианта, необходимо вычислить коэффициенты разделяющей линии, которая определяется выражением:
2. Вычислить отступы (margin) для зеленых точек (с меткой класса -1) и синих точек (с меткой класса +1). Напомню, что отступ вычисляется по формуле:

где - метка класса образа (точки) ; - скалярное произведение векторов.

Вектор должен быть подобран так, чтобы для наиболее удаленных точек от разделяющей линии отступ был положительным, а для ближних - отрицательным.

Рис. 1. График для 5 варианта

# Краткая теория

Прогноз линейной регрессионной модели определяется уравнением:

где – спрогнозированное значение, – количество признаков, – значение -того признака, – -тый параметр модели (включая член смещения и веса признаков ).

Обучение модели означает установку её параметров так, чтобы модель была наилучшим образом подогнана к обучающему набору. Для этой цели первым делом нужна мера того, насколько хорошо (или плохо) модель подогнана к обучающим данным. Самой распространенной мерой производительности регрессионной модели является квадратный корень из среднеквадратической ошибки (RMSE). Однако в данной лабораторной работе не стоит задача определения наилучших параметров, поэтому мы не будем вычислять RMSE.

Подбор гиперпараметров будет осуществляться с использованием ступенчатой функции Хевисайда, хорошо подходящей для задач бинарной классификации:

На Рис. 2 для модели хорошо видно, что радиус-векторы точек для класса образуют острый угол с вектором . Т.к. скалярное произведение – это, фактически, вычисление косинуса угла между векторами, то для острых углов будут получены положительные значения. Точки же класса образуют тупые углы с вектором . Следовательно, для будут получены отрицательные значения.

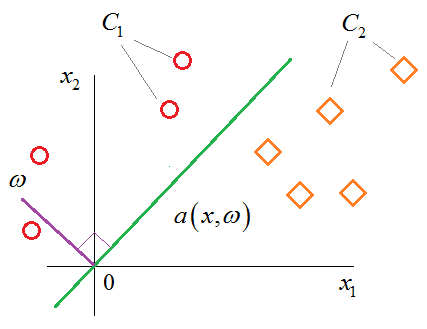


Рис. 2. Демонстрация принципа классификации функцией Хевисайда

В итоге, алгоритм классификации образов с помощью модели можно записать в виде:

Здесь – знаковая функция, которая возвращает +1 для положительных чисел и -1 – для отрицательных:

Математически расчет оптимальной линии разделение осуществляется по минимизации функционала качества Q:



(1)

, где M – отступ точки от линии разделения

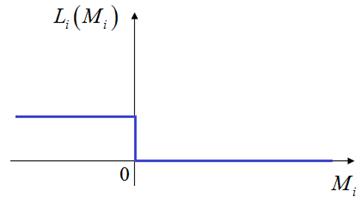
(2)

В качестве функции потерь используется кусочно-непрерывная функция L

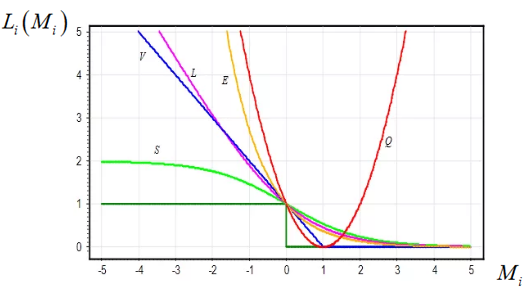
(3)

Здесь квадратные скобки – это индикатор ошибки. В соответствии с нотацией Айверсона они возвращают 1 для True и 0 – для False.

Если изобразить эту функцию потерь, зависящую от отступа, то получим ступенчатый график:

Рисунок 3. Кусочная функция ошибки

Недостаток этой функции в том что она не имеет производных в точках перегиба, а значит аналитическое решение минимизации значения найти невозможно, только перебором. Однако в математике существует много похожих по свойствам (точка экстремума в 0, 1) дифференцируемых функций. Вот некоторые, наиболее известные:

Рисунок 4. Виды функций ошибок

Для нашей задачи подойдет типовая квадратичная функция (FLD) .

(4)

После подстановки в (1) получим функционал качества:



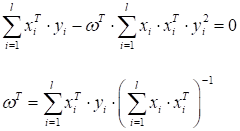
(5)

Тогда оптимальный вектор ω можем найти как производную функционала качества.



(6)

Откуда



# Результаты работы

Работа была выполнена на языке программирования Python 3.

Результат работы программы приведен на Рис. 3.

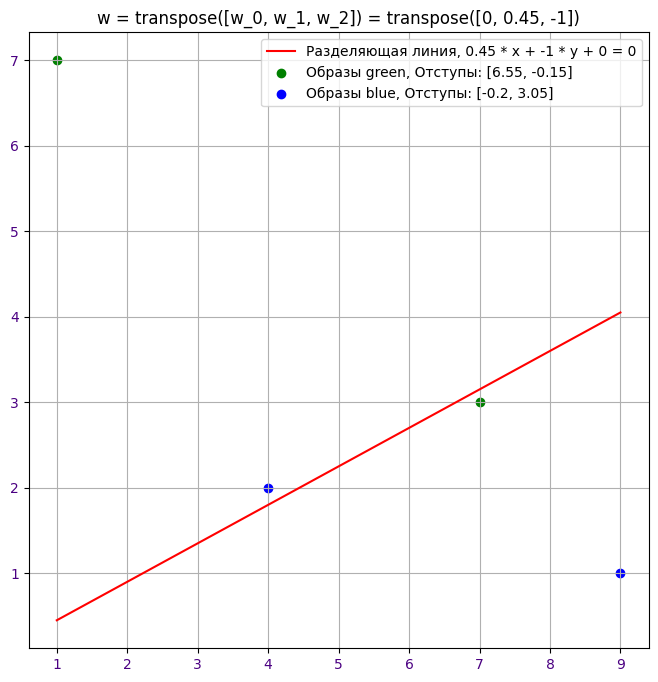


Рис. 3. Полученный график

# Вывод

В ходе лабораторной работы я определил коэффициенты разделяющей линии и отступы для зеленых и синих точек.