

## **Постановка задачи по реализации алгоритма логистической регрессии**

### **1. Логистическая регрессия**

В этой части задания необходимо построить модель логистической регрессии, чтобы предсказать, поступит ли студент в университет. Предположим, вы хотите определить шансы каждого абитуриента на поступление на основе его баллов на двух экзаменах. В файле ex2data1.txt содержатся данные предыдущих кандидатов, которые вы можете использовать в качестве обучающего набора для логистической регрессии. Первые два столбца – баллы абитуриента по двум экзаменам, третий столбец – решение о зачислении.

Ваша задача — построить классификационную модель, которая оценивает вероятность поступления абитуриента на основе результатов этих двух экзаменов.

- 1.1. Загрузить данные из файла ex2data1.txt. Визуализировать данные.
- 1.2. Реализовать функцию потерь (желательно в виде отдельной функции).
- 1.3. Реализовать градиентный спуск (желательно в виде отдельной функции). Задать скорость обучения равной 0.01, начальные значения параметров выбрать нулевыми.
- 1.4. Проверить, что алгоритм обучения сходится, построив кривую зависимости значения функции потерь от номера итерации обучения. Попробовать выбрать другую скорость обучения для увеличения скорости сходимости к минимуму функции потерь, отметить, как это отразилось на кривой обучения.
- 1.5. Построить границу принятия решения
- 1.6. Предсказать, с какой вероятностью поступит абитуриент, набравший 45 баллов по первому экзамену и 85 баллов по второму.
- 1.7. Посчитать точность классификации на обучающей выборке:  
$$\text{точность} = (\text{количество правильно классифицированных примеров}) / (\text{размер обучающей выборки}).$$

## 2. Логистическая регрессия с регуляризацией

В этой части задания необходимо реализовать регуляризованную логистическую регрессию с нелинейной границей решения, чтобы предсказать, проходят ли микрочипы, произведенные на заводе-изготовителе, систему контроля качества. В ходе контроля качества каждый микрочип проходит различные тесты, чтобы убедиться в его правильном функционировании. На основе двух тестов вы хотите определить, соответствует ли микрочип характеристикам качества. В файле `ex2data2.txt` содержатся данные о результатах испытаний прошлых микрочипов, которые вы можете использовать в качестве обучающего набора для логистической регрессии. Первые два столбца – оценка микрочипа по двум тестам, третий столбец – решение о соответствии допустимым характеристикам.

Ваша задача — построить классификационную модель, которая оценивает вероятность того, что конкретный микрочип прошел контроль качества.

2.1. Загрузить данные из файла `ex2data2.txt`. Визуализировать данные.

2.2. Создать новый вектор признаков для каждого элемента обучающей выборки:  $\vec{x'}^T = [1, x_1, x_2, x_1^2, x_2^2, x_1 \cdot x_2, \dots, x_1^{30}, x_2^{30}]$  размерности 496 (т.е.  $\vec{\theta}^T \vec{x'}$  – полином 30-й степени).

2.3 Разбить данные на тренировочную и тестовую выборки в соотношении 80/20.

2.3. Реализовать регуляризованную функцию потерь.

2.4. Реализовать градиентный спуск с регуляризацией.

2.5. Задать скорость обучения равной 0.5, начальные значения параметров выбрать нулевыми, параметр регуляризации а) 0.0, б) 1.0. Обучить модель с так заданными параметрами, посчитать точность на тренировочной и тестовой выборках. Сравнить, как изменились точности с регуляризацией и без нее.

2.6. Для двух вариантов параметра регуляризации построить границу принятия решения.