Министерство образования Республики Беларусь

Учреждение образования

БелорусскиЙ государственный университет

информатики и радиоэлектроники

Факультет компьютерных систем и сетей

Кафедра информатики

ОТЧЕТ

по лабораторной работе №2

**Логистическая регрессия. Многоклассовая классификация**

|  |  |  |
| --- | --- | --- |
| Студент |  | И.Ю. Фалько |
| Преподаватель |  | М. В. Стержанов |

Минск 2019

СОДЕРЖАНИЕ

[Условие лабораторной работы 3](#_Toc25711884)

[Ход выполнения 3](#_Toc25711885)

[Заключение 10](#_Toc25711886)

**Условие лабораторной работы**

Набор данных **ex2data1.txt** представляет собой текстовый файл, содержащий информацию об оценке студента по первому экзамену (первое число в строке), оценке по второму экзамену (второе число в строке) и поступлении в университет (0 - не поступил, 1 - поступил).

Набор данных **ex2data2.txt** представляет собой текстовый файл, содержащий информацию о результате первого теста (первое число в строке) и результате второго теста (второе число в строке) изделий и результате прохождения контроля (0 - контроль не пройден, 1 - контроль пройден).

Набор данных **ex2data3.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит 5000 изображений 20x20 в оттенках серого. Каждый пиксель представляет собой значение яркости (вещественное число). Каждое изображение сохранено в виде вектора из 400 элементов. В результате загрузки набора данных должна быть получена матрица 5000x400. Далее расположены метки классов изображений от 1 до 9 (соответствуют цифрам от 1 до 9), а также 10 (соответствует цифре 0).

# Ход выполнения

1. Загрузите данные **ex2data1.txt** из текстового файла.
2. Постройте график, где по осям откладываются оценки по предметам, а точки обозначаются двумя разными маркерами в зависимости от того, поступил ли данный студент в университет или нет.

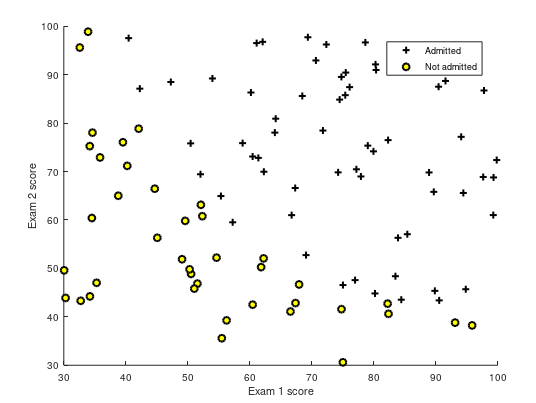
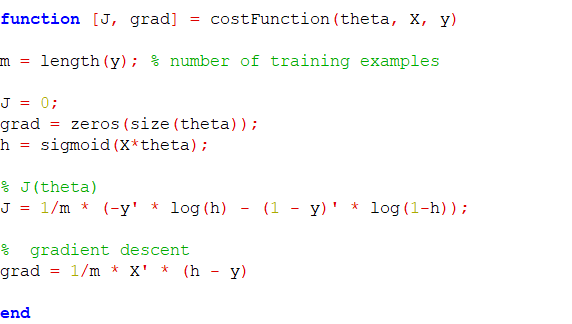


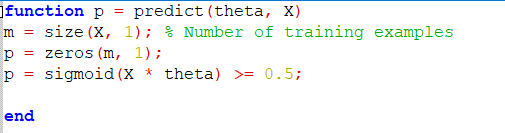
Рисунок 1 – график зависимости поступления от оценок по предметам

1. Реализуйте функции потерь J(θ) и градиентного спуска для логистической регрессии с использованием векторизации.

ы

1. Реализуйте функцию предсказания вероятности поступления студента в зависимости от значений оценок по экзаменам.

Функция предсказания:



1. Постройте разделяющую прямую, полученную в результате обучения модели. Совместите прямую с графиком из пункта 2.

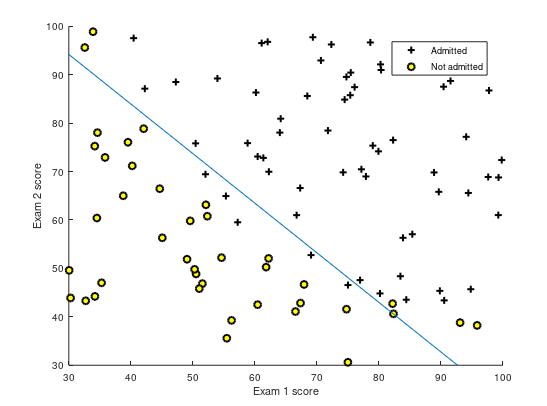


Рисунок 3 – оценки студентов и разделяющая прямая

1. Загрузите данные **ex2data2.txt** из текстового файла.
2. Постройте график, где по осям откладываются результаты тестов, а точки обозначаются двумя разными маркерами в зависимости от того, прошло ли изделие контроль или нет.

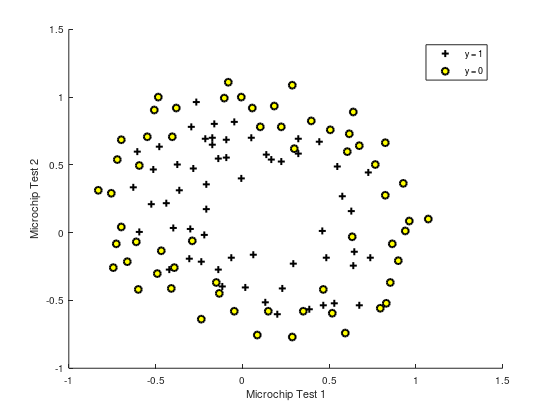
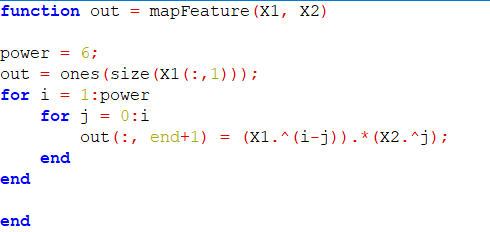
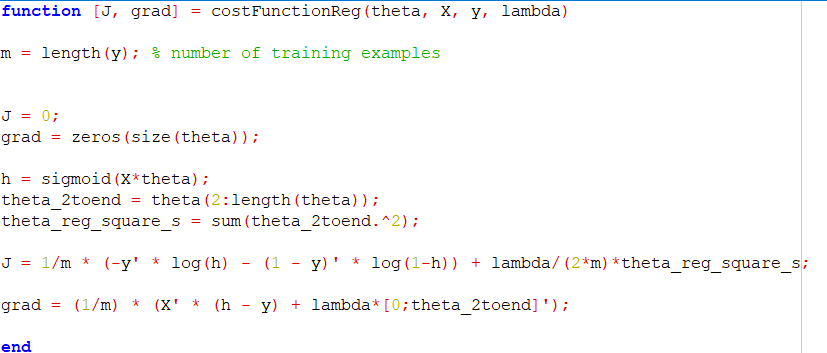


Рисунок 4 – результаты тестов

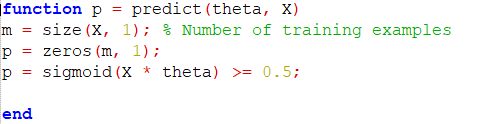
1. Постройте все возможные комбинации признаков x1 (результат первого теста) и x2 (результат второго теста), в которых степень полинома не превышает 6, т.е. 1, x1, x2, x12, x1x2, x22, …, x1x25, x26 (всего 28 комбинаций).



1. Реализуйте L2-регуляризацию для логистической регрессии и обучите ее на расширенном наборе признаков методом градиентного спуска.



1. Реализуйте функцию предсказания вероятности прохождения контроля изделием в зависимости от результатов тестов.



1. Постройте разделяющую кривую, полученную в результате обучения модели. Совместите прямую с графиком из пункта 7.

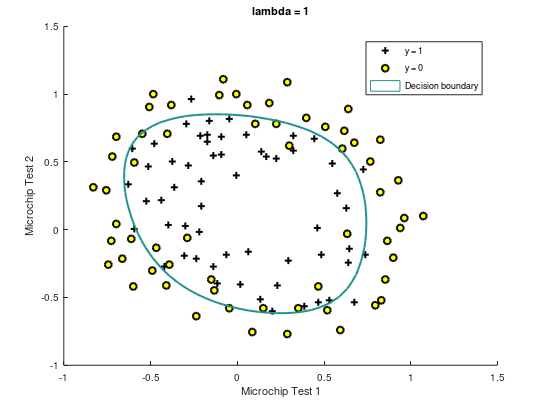


Рисунок 5 – разделяющая прямая

1. Попробуйте различные значения параметра регуляризации λ. Как выбор данного значения влияет на вид разделяющей кривой? Ответ дайте в виде графиков.

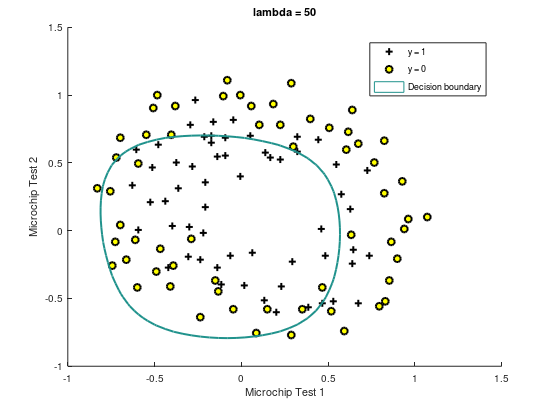


Рисунок 6 – лямбда равна 50

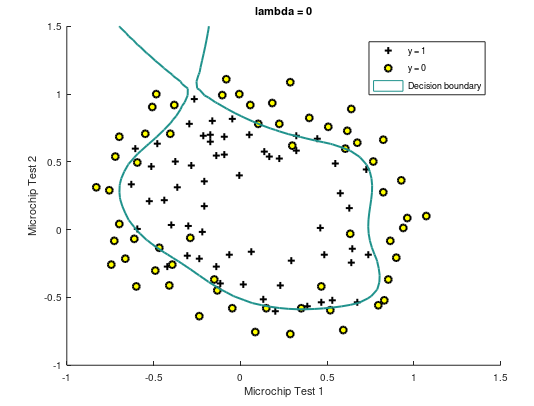


Рисунок 7 – лямбда равна 0

1. Загрузите данные **ex2data3.mat** из файла.
2. Визуализируйте несколько случайных изображений из набора данных. Визуализация должна содержать каждую цифру как минимум один раз.

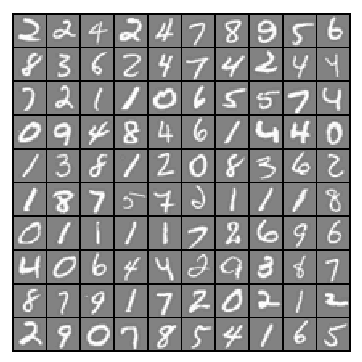
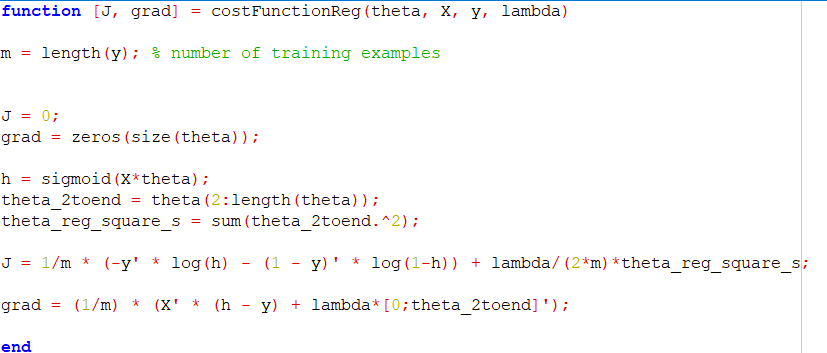
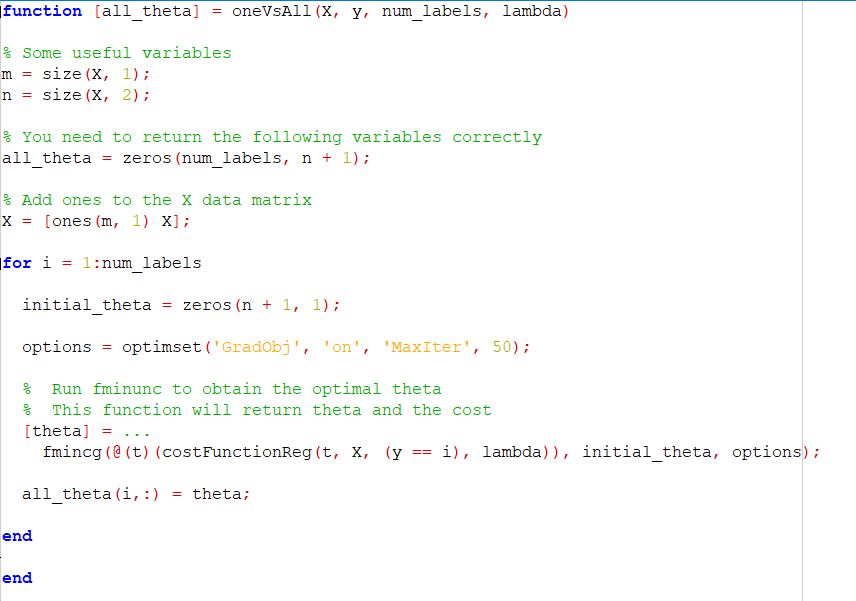


Рисунок 7 – визуализация данных

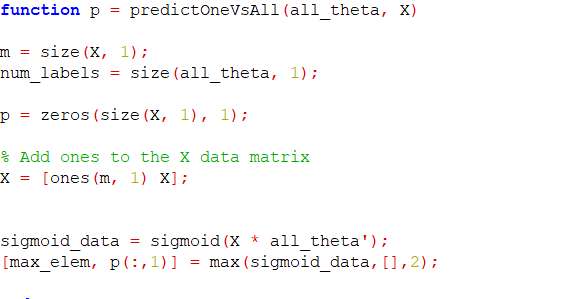
1. Реализуйте бинарный классификатор с помощью логистической регрессии с использованием векторизации (функции потерь и градиентного спуска).
2. Добавьте L2-регуляризацию к модели.



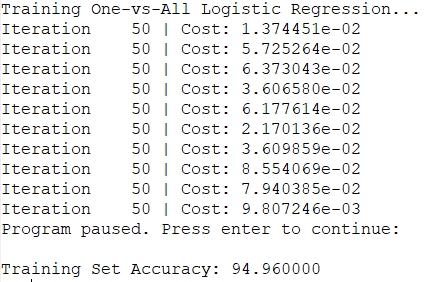
1. Реализуйте многоклассовую классификацию по методу “один против всех”.



1. Реализуйте функцию предсказания класса по изображению с использованием обученных классификаторов.



1. Процент правильных классификаций на обучающей выборке должен составлять около 95%.



# ЗАКЛЮЧЕНИЕ

В данной лабораторной работе были освоены навыки логической регрессии с помощью градиентного спуска, нахождение разделяющей границы. Также было использован метод «один против всех» для многоклассовой классификации.

Была изучена проблема переобучения алгоритма. Для того, чтобы его избежать, необходимо уменьшить кол-во признаков, либо провести регуляризацию.