|  |  |  |  |
| --- | --- | --- | --- |
| Министерство образования Республики Беларусь  Учреждение образования  Белорусский Государственный Университет Информатики и Радиоэлектроники | | | |
| Факультет | Компьютерных сетей и систем | | |
| Кафедра | Информатики  Дисциплина: Конструирование те технологии электронных вычислительных средств | | |
|  |  | | |
| **лабораторная работа №2**  по курсу Машинное обучение  на тему  **ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ. МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ** | | | |
| Магистрант:  гр. 858341  Кудин Н.И. | |  | Проверил:  Стержанов М.В. |
| Минск, 2019 | | | |

**Постановка задачи**

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 работы**

1. Загрузим данные ex2data1.txt из текстового файла

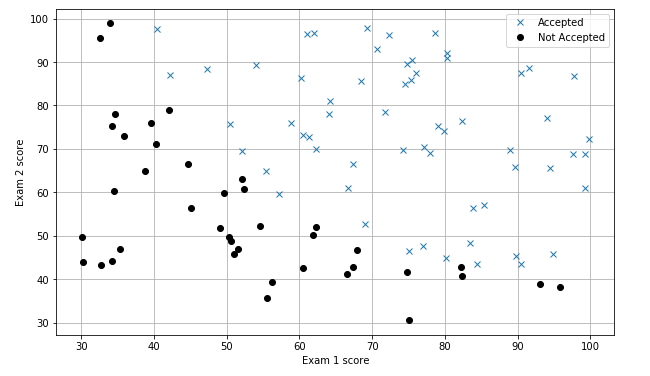
*datafile = "ex2data1.txt"*

*#!head $datafile*

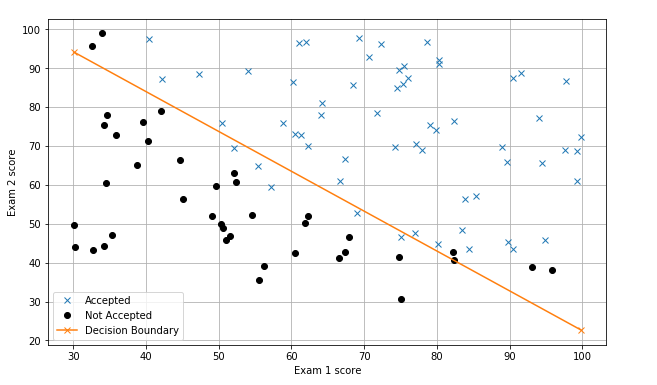
*# read first file with student results*

*marksData = np.loadtxt(datafile,delimiter=',',usecols=(0,1,2),unpack=True)*

1. Построим 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.



1. Построим разделяющую прямую, полученную в результате обучения модели. Совместим прямую с графиком из пункта 1.



1. Построение графика на основе данных из ex2data2.txt

*datafile = "ex2data2.txt"*

*columns = np.loadtxt(datafile,delimiter=',',usecols=(0,1,2),unpack=True)*

*x = np.transpose(np.array(columns[:-1]))*

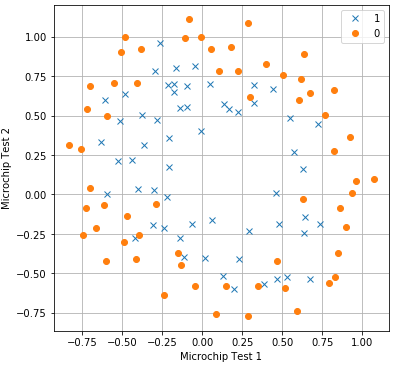
*y = np.transpose(np.array(columns[-1:]))*

*m = y.size # number of training examples*

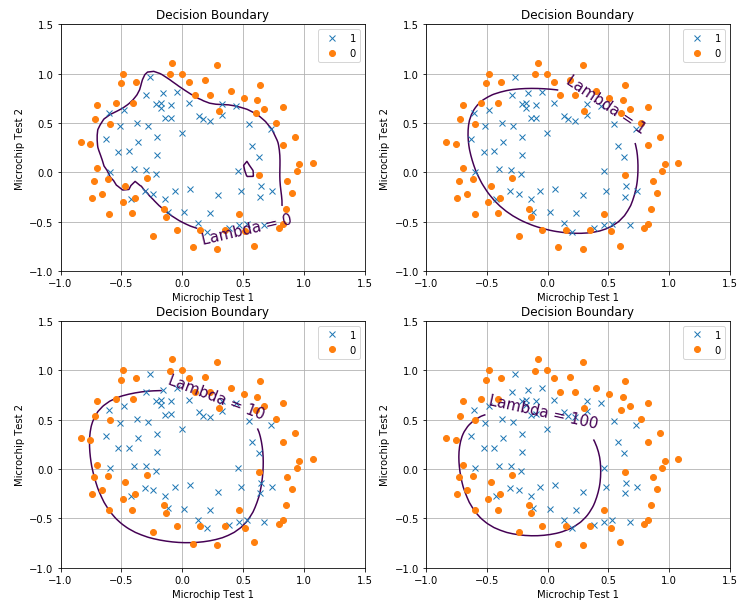
*# insert the usual column of one into the "x" matrix*

*x = np.insert(x,0,1,axis=1)*

1. Построим график, где по осям откладываются результаты тестов, а точки обозначаются двумя разными маркерами в зависимости от того, прошло ли изделие контроль или нет.

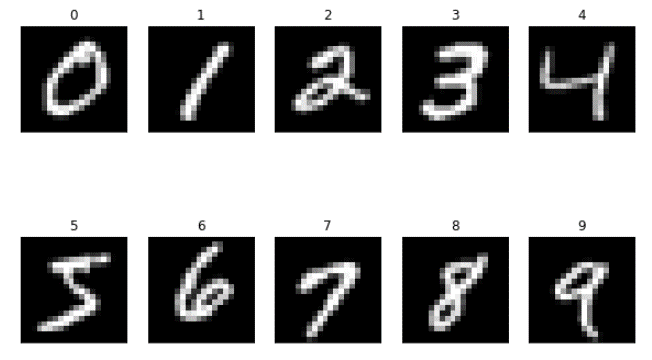


1. Построим разделяющую кривую, полученную в результате обучения модели. Совместим прямую с графиком из пункта выше.

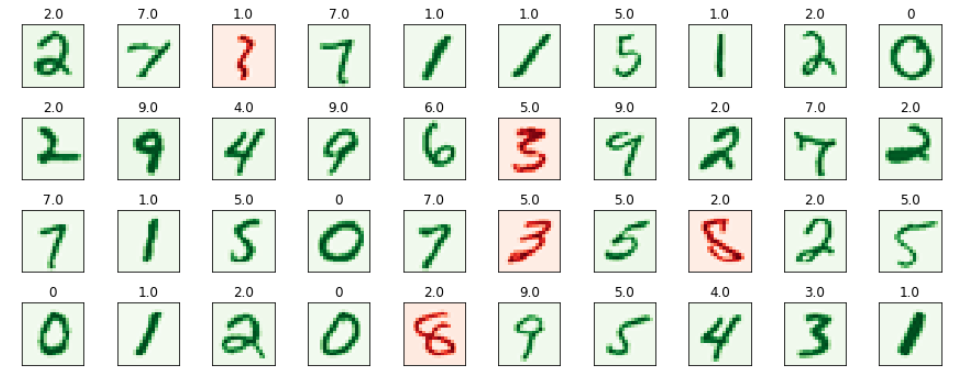


Как видно из графиков, при значении регуляризации = 0 наша модель переобучилась, а при значении = 100 недообучилась.

1. Визуализируем несколько случайных изображений из набора данных. Визуализация содержит каждую цифру как минимум один раз.



1. Реализуем многоклассовую классификацию по методу “один против всех”. Кроме того, реализуем функцию предсказания класса по изображению с использованием обученных классификаторов. Процент правильных классификаций на обучающей выборке должен составлять около 95%.



**Заключение**

В ходе лабораторной работы были изучены такие программные средства как Jupiter, библиотеки matplotlib, pandas, scipy и numpy.

Опытным путем были получены результаты распознавания изображений с помощью многоклассовой классификации по методу “один против всех”

Логистическая регрессия полезна для ситуаций, в которых необходимо иметь возможность предсказать наличие или отсутствие характеристики или итога на основании значений набора переменных – предикторов. Обладает некоторыми недостатками, свойственными линейной регрессии – низкая устойчивость к ошибкам, зависимость от набора данных, но в общем работает лучше, и может быть приведена к виду линейной регрессии для упрощения вычислений.