Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра информатики

Дисциплина: Машинное обучение (МО)

Отчет о лабораторной работе №1

«Линейная регрессия»

|  |
| --- |
| **Выполнил:** |
| **Студент гр. 858341** |
| **Немкович А. В.** |

|  |
| --- |
| **Проверил:** |
| **Стержанов М. В.** |

**Минск 2019**

Содержание

[Постановка задачи 3](#_Toc25165821)

[Ход выполнения 4](#_Toc25165822)

[Выводы 9](#_Toc25165823)

# Постановка задачи

Набор данных **ex1data1.txt** представляет собой текстовый файл, содержащий информацию о населении городов (первое число в строке) и прибыли ресторана, достигнутой в этом городе (второе число в строке). Отрицательное значение прибыли означает, что в данном городе ресторан терпит убытки.

Набор данных **ex1data2.txt** представляет собой текстовый файл, содержащий информацию о площади дома в квадратных футах (первое число в строке), количестве комнат в доме (второе число в строке) и стоимости дома (третье число).

**Задание.**

1. Загрузите набор данных **ex1data1.txt** из текстового файла.
2. Постройте график зависимости прибыли ресторана от населения города, в котором он расположен.
3. Реализуйте функцию потерь J(θ) для набора данных **ex1data1.txt**.
4. Реализуйте функцию градиентного спуска для выбора параметров модели. Постройте полученную модель (функцию) совместно с графиком из пункта 2.
5. Постройте трехмерный график зависимости функции потерь от параметров модели (θ0 и θ1) как в виде поверхности, так и в виде изолиний (contour plot).
6. Загрузите набор данных **ex1data2.txt** из текстового файла.
7. Произведите нормализацию признаков. Повлияло ли это на скорость сходимости градиентного спуска? Ответ дайте в виде графика.
8. Реализуйте функции потерь J(θ) и градиентного спуска для случая многомерной линейной регрессии с использованием векторизации.
9. Покажите, что векторизация дает прирост производительности.
10. Попробуйте изменить параметр ɑ (коэффициент обучения). Как при этом изменяется график функции потерь в зависимости от числа итераций градиентного спуск? Результат изобразите в качестве графика.
11. Постройте модель, используя аналитическое решение, которое может быть получено методом наименьших квадратов. Сравните результаты данной модели с моделью, полученной с помощью градиентного спуска.
12. Ответы на вопросы представьте в виде отчета.

# Ход выполнения

**Задания 1-2**: загрузка данных и построение графика зависимости прибыли ресторана от населения города.

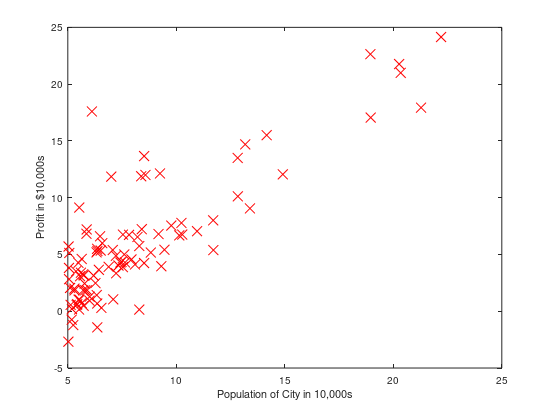
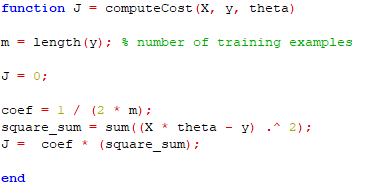


Рис 1. – график зависимости прибыли ресторана от населения города

**Задание 3**: реализация функции потерь J(θ).

Код функции потерь:



**Задание 4**: реализация функции градиентного спуска и построение модели совместно с графиком из пункта 2.

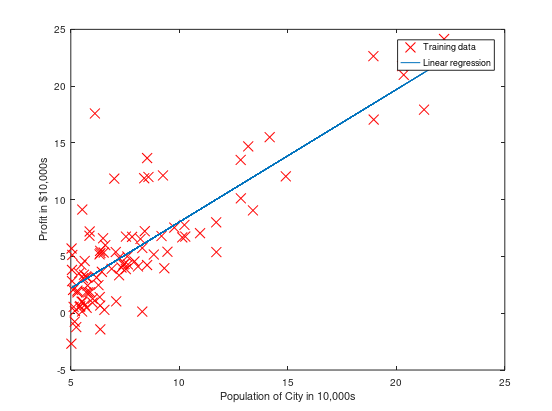


Рис 2. – график модели совместно с исходными данными

**Задание 5**: построение трехмерного графика зависимости функции потерь от параметров модели (θ0 и θ1) как в виде поверхности, так и в виде изолиний

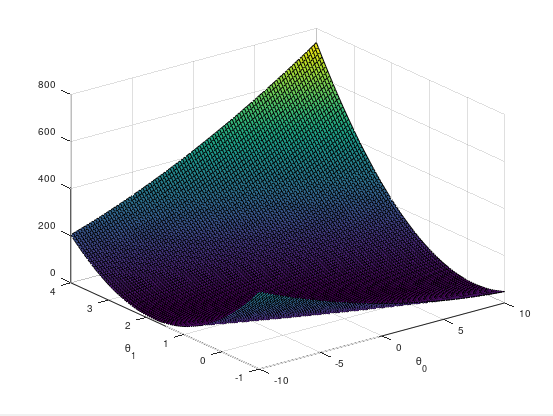


Рис 3. – график зависимости функции потерь от параметров модели в виде поверхности

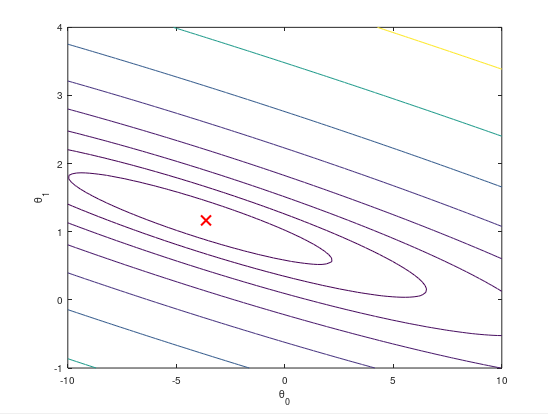


Рис 4. – график зависимости функции потерь от параметров модели в виде изолиний

**Задания 6-7**: загрузка данных и построение графика сходимости градиентного спуска

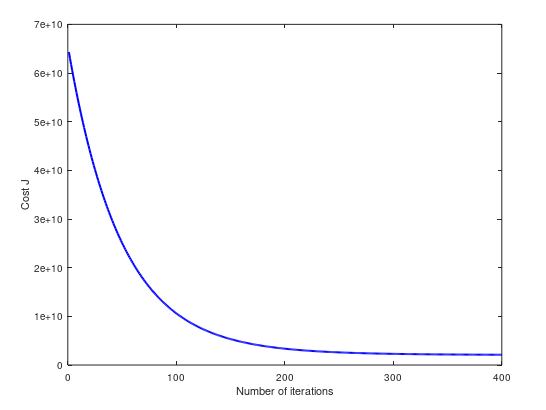
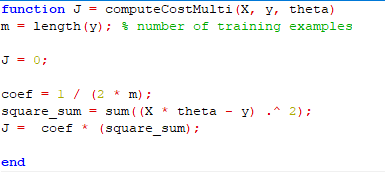


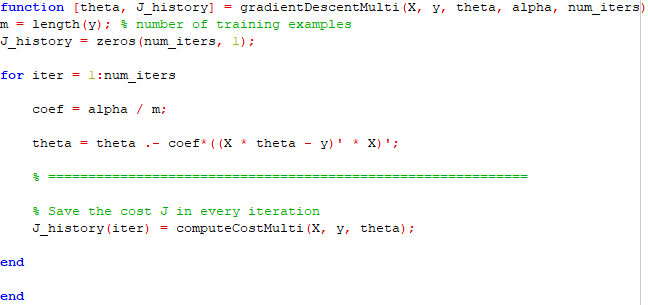
Рис 5. – график сходимости градиентного спуска

**Задания 8-9**: реализация функции потерь J(θ) и градиентного спуска для случая многомерной линейной регрессии с использованием векторизации

Код функции потерь:



Код функции градиентного спуска:



**Задание 10**: реализация графика функции потерь в зависимости от числа итераций градиентного спуска с разными значениями коэффициента обучения

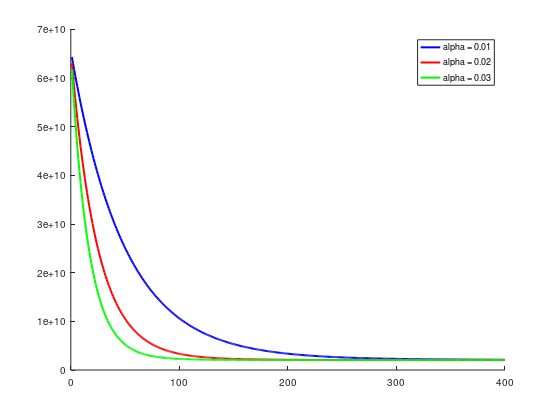
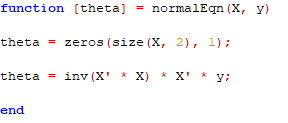


Рис 6. – график функции потерь в зависимости от числа итераций градиентного спуска

**Задание 11**: реализация модели, используя аналитическое решение. И сравнение результатов с градиентным спуском

Для этого было реализован метод Нормальных уравнений (Normal Equations)

Код данного метода:



Сравнение результатов:

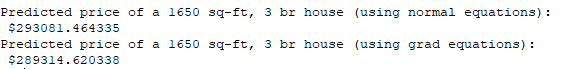


Рис 7. – Сравнение результатов двух моделей

# Выводы

Мы научились решать задачу линейной регрессии с помощью градиентного спуска. Также мы изучили и аналитическое решение для нахождения минимального значения функции потерь. Данные два метода имеют свои достоинства и недостатки. Сравнительная характеристика этих методов представлена следующей таблицей.

Таблица 1. Сравнение градиентного спуска и метода Нормальных уравнений

|  |  |
| --- | --- |
| **Градиентный спуск** | **Нормальное уравнение** |
| Необходимо выбрать коэффициент обучения | Нет необходимости выбирать коэффициент обучения |
| Много итераций | Не итерируемый |
|  | , необходимая для расчета |
| Отлично работает когда *n* - большое | Плохо работает если *n* - большое |