Министерство образования Республики Беларусь

Учреждение образования

БелорусскиЙ государственный университет

информатики и радиоэлектроники

Факультет компьютерных систем и сетей

Кафедра информатики

ОТЧЕТ

по лабораторной работе №3

**Переобучение и регуляризация**

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

Минск 2019

СОДЕРЖАНИЕ

[Условие лабораторной работы 3](#_Toc25711884)

[Ход выполнения 3](#_Toc25711885)

[Заключение 11](#_Toc25711886)

**Условие лабораторной работы**

Набор данных **ex3data1.mat** представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит две переменные X (изменения уровня воды) и y (объем воды, вытекающий из дамбы). По переменной X необходимо предсказать y. Данные разделены на три выборки: обучающая выборка (X, y), по которой определяются параметры модели; валидационная выборка (Xval, yval), на которой настраивается коэффициент регуляризации; контрольная выборка (Xtest, ytest), на которой оценивается качество построенной модели.

# Ход выполнения

1. Загрузите данные **ex3data1.mat** из файла.
2. Постройте график, где по осям откладываются X и y из обучающей выборки.

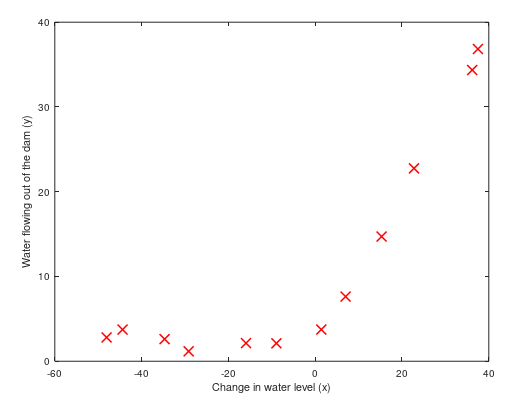
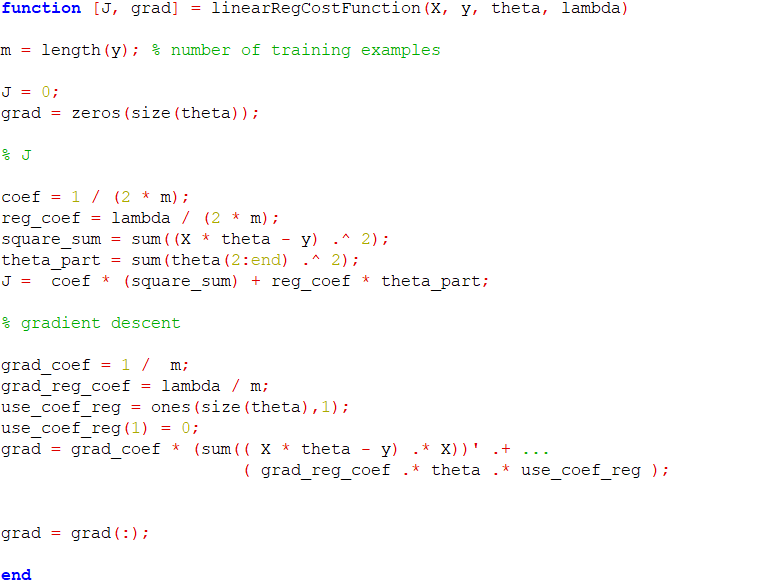


Рисунок 1 – данные из файла

1. Реализуйте функцию стоимости потерь для линейной регрессии с L2-регуляризацией.
2. Реализуйте функцию градиентного спуска для линейной регрессии с L2-регуляризацией.



1. Постройте модель линейной регрессии с коэффициентом регуляризации 0 и постройте график полученной функции совместно с графиком из пункта

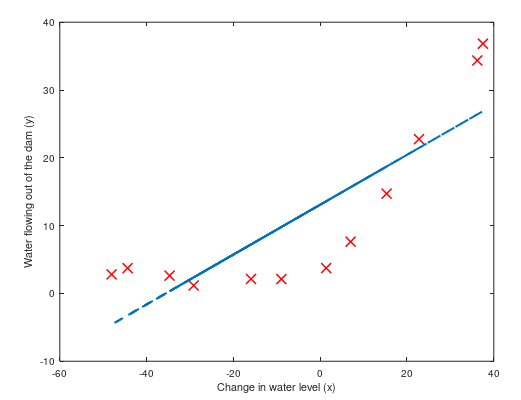
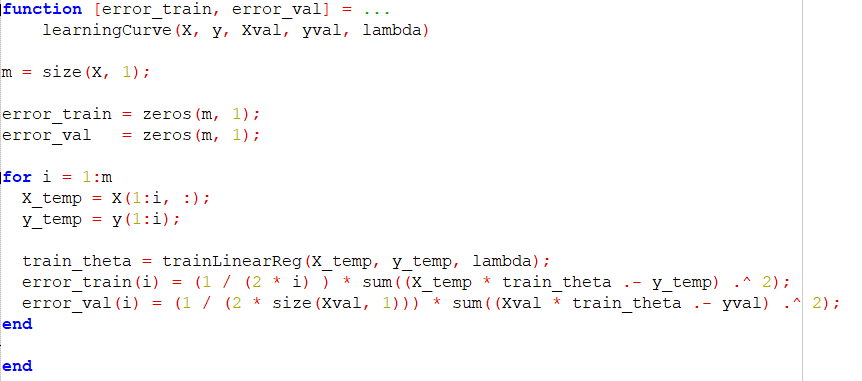


Рисунок 2 – модель линейной регрессии

1. Постройте график процесса обучения (learning curves) для обучающей и валидационной выборки. По оси абсцисс откладывается число элементов из обучающей выборки, а по оси ординат - ошибка (значение функции потерь) для обучающей выборки (первая кривая) и валидационной выборки (вторая кривая).



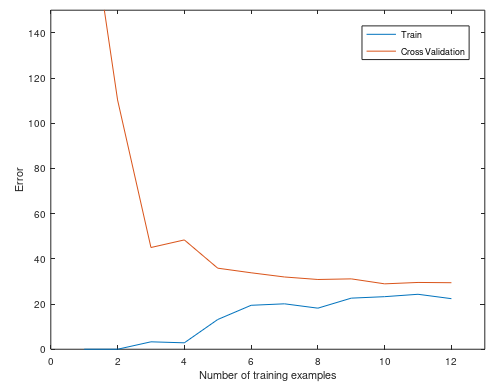
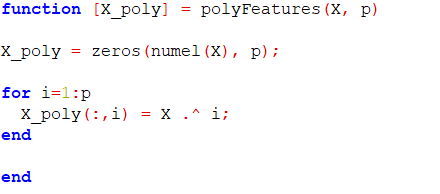
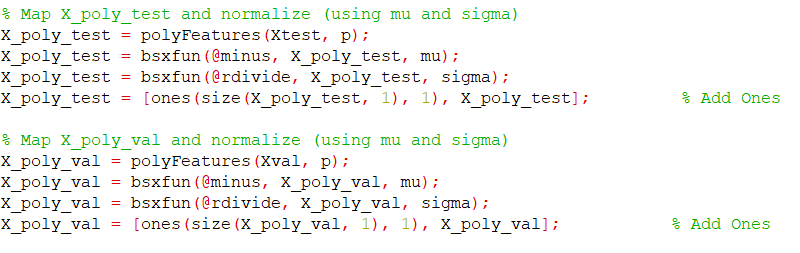


Рисунок 3 – график процесса обучения

1. Реализуйте функцию добавления p - 1 новых признаков в обучающую выборку (X2, X3, X4, …, Xp).



1. Поскольку в данной задаче будет использован полином высокой степени, то необходимо перед обучением произвести нормализацию признаков.



1. Обучите модель с коэффициентом регуляризации 0 и p = 8.
2. Постройте график модели, совмещенный с обучающей выборкой, а также график процесса обучения.

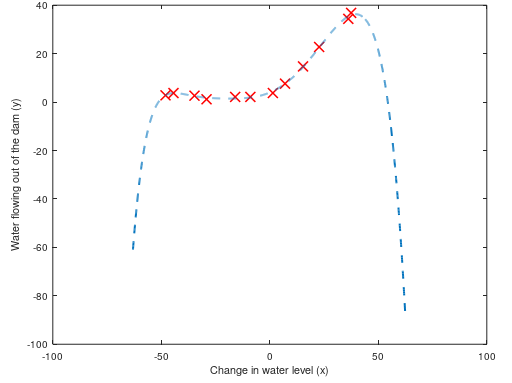


Рисунок 4 – график с обучающей выборкой

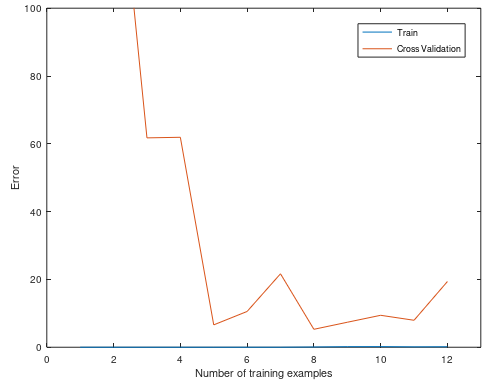


Рисунок 5 – процесс обучения

1. Постройте графики из пункта 10 для моделей с коэффициентами регуляризации 1 и 100.

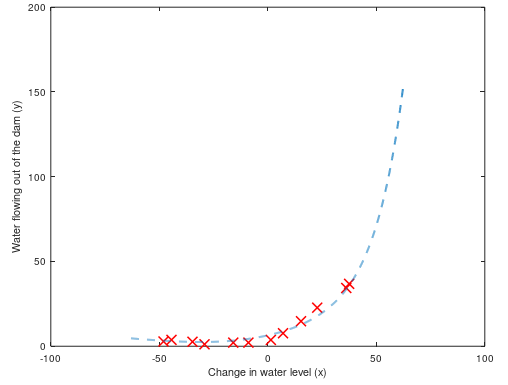


Рисунок 6 –модель, совмещенная с обучающей выборкой при лямбда = 1

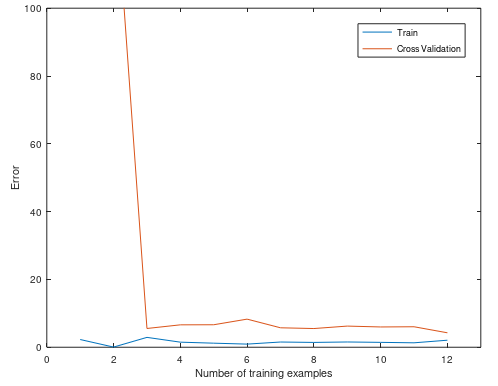


Рисунок 7 – процесс обучения при лямбда = 1

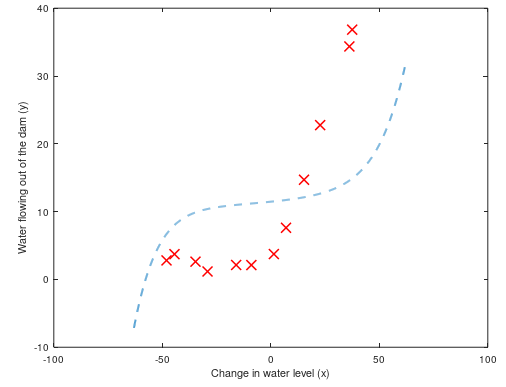


Рисунок 8 – модель, совмещенная с обучающей выборкой при лямбда = 100

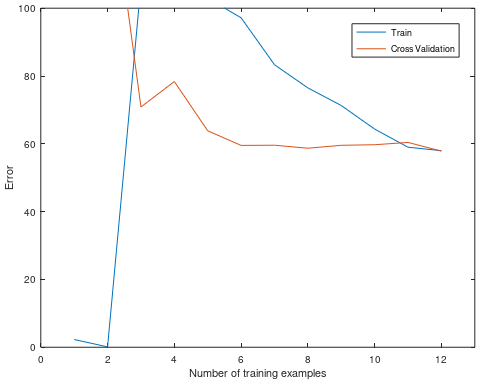
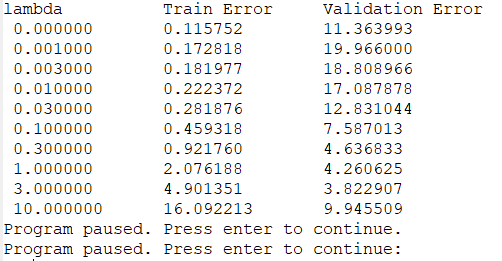


Рисунок 9 – процесс обучения при лямбда = 100

1. С помощью валидационной выборки подберите коэффиент регуляризации, который позволяет достичь наименьшей ошибки. Процесс подбора отразите с помощью графика (графиков).



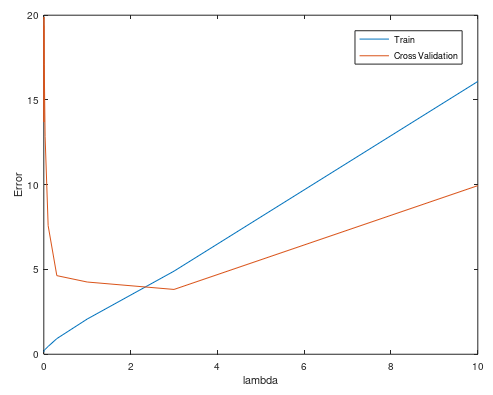


Рисунок 10– график подбора коэффициента регуляризации

1. Вычислите ошибку (потерю) на контрольной выборке.



# ЗАКЛЮЧЕНИЕ

В данной лабораторной работе были освоены навыки определять проблемы недообучения и переобучения модели с помощью графика, выявлять и предотвращать переобучение алгоритма.

В ходе изучения проблемы переобучения алгоритма было усвоено, что для того, чтобы избежать проблема переобучения, нужно увеличить кол-во обучающих примеров, увеличить коэффициент регуляризации лямбда, либо уменьшить набор признаков. Для того, чтобы его избежать недообучения, необходимо добавить полиноминальные признаки, уменьшить лямбда.