Министерство образования Республики Беларусь

Учреждение образования

"Белорусский Государственный университет информатики

и радиоэлектроники"

Лабораторная работа №3

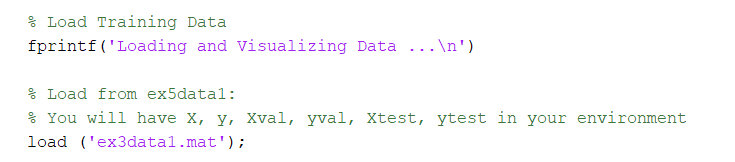
по учебной дисциплине “Машинное обучение”

|  |  |
| --- | --- |
| Выполнил: | Студент гр. 956241 Дубовик Н.О. |
|  |  |

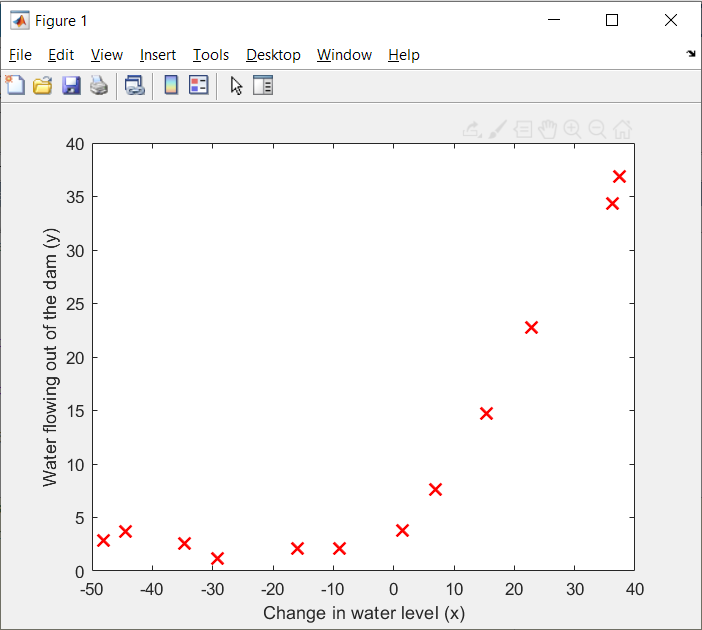
Минск 2019

**Задание.**

Загрузите данные **ex3data1.mat** из файла.

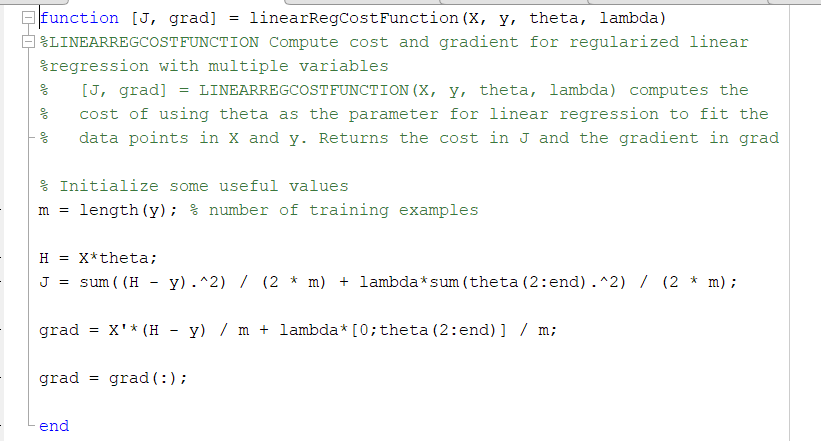


Постройте график, где по осям откладываются X и y из обучающей выборки.



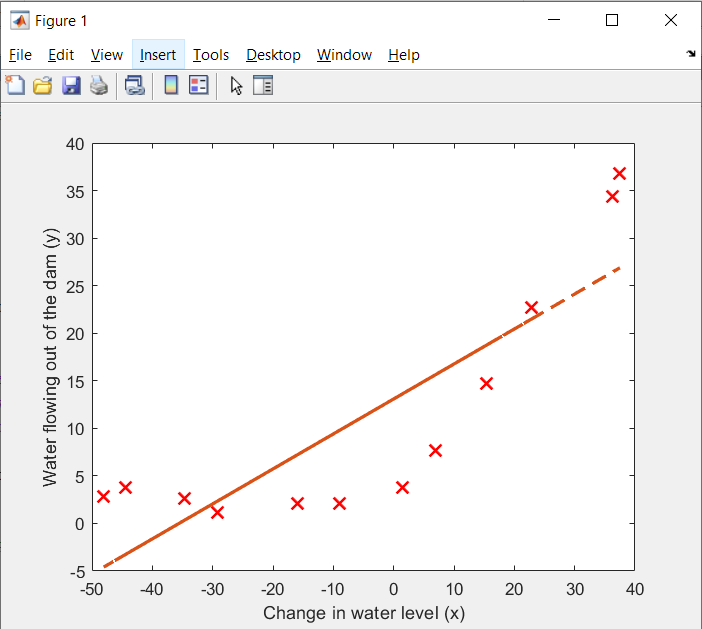
Реализуйте функцию стоимости потерь для линейной регрессии с L2-регуляризацией.

Реализуйте функцию градиентного спуска для линейной регрессии с L2-регуляризацией.

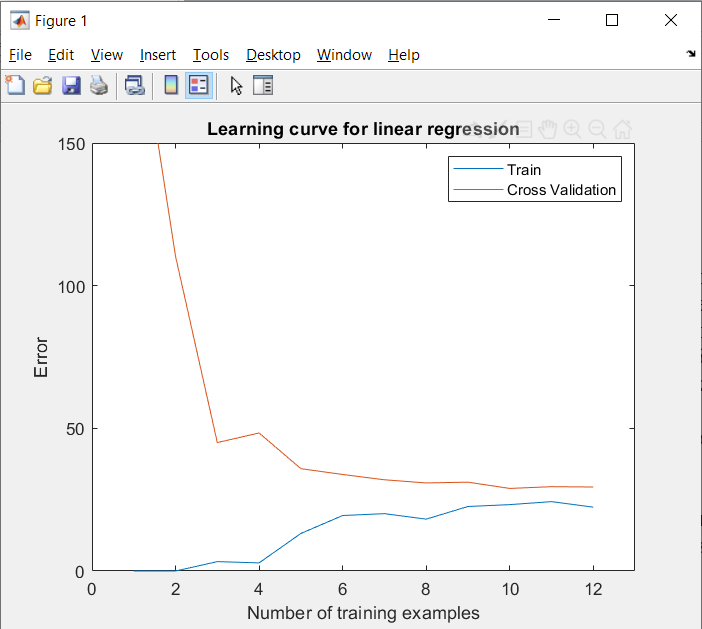


Постройте модель линейной регрессии с коэффициентом регуляризации 0 и постройте график полученной функции совместно с графиком из пункта 2. Почему регуляризация в данном случае не сработает?

Регуляризация не сработает из-за низкой размерности.

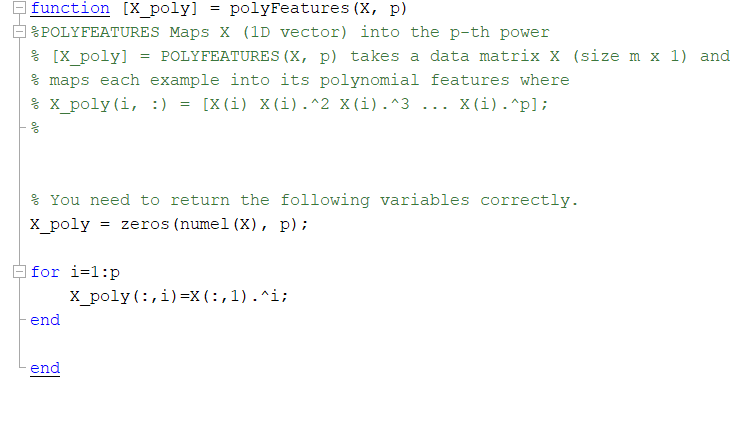


Постройте график процесса обучения (learning curves) для обучающей и валидационной выборки. По оси абсцисс откладывается число элементов из обучающей выборки, а по оси ординат - ошибка (значение функции потерь) для обучающей выборки (первая кривая) и валидационной выборки (вторая кривая). Какой вывод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

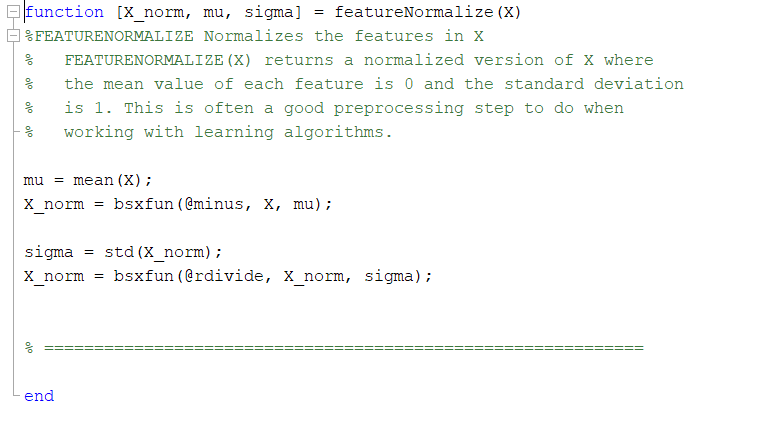


Данные кривые показывают проблему большого смещения, так как модель линейной регрессии слишком просто и не может хорошо соответствовать набору данных.

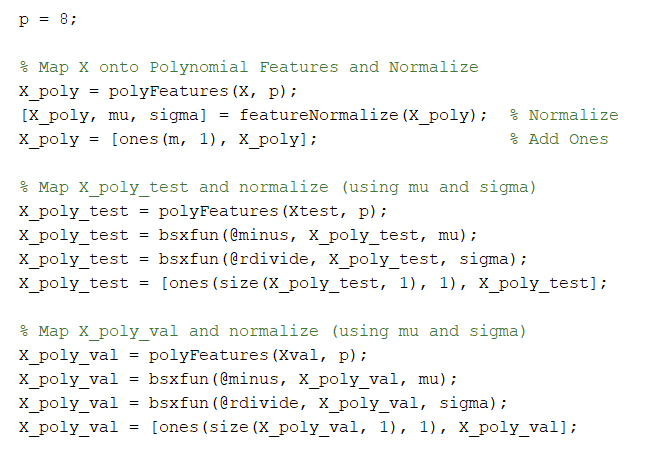
Реализуйте функцию добавления p - 1 новых признаков в обучающую выборку (X2, X3, X4, …, Xp).



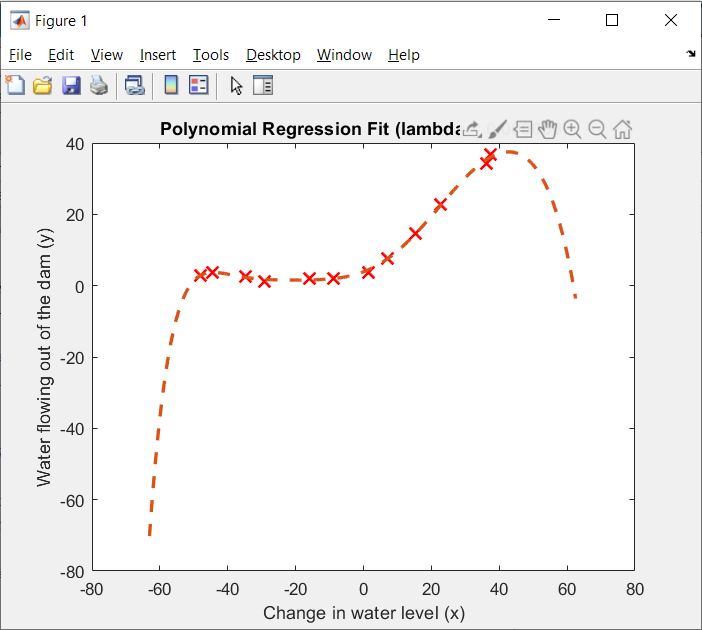
Поскольку в данной задаче будет использован полином высокой степени, то необходимо перед обучением произвести нормализацию признаков.

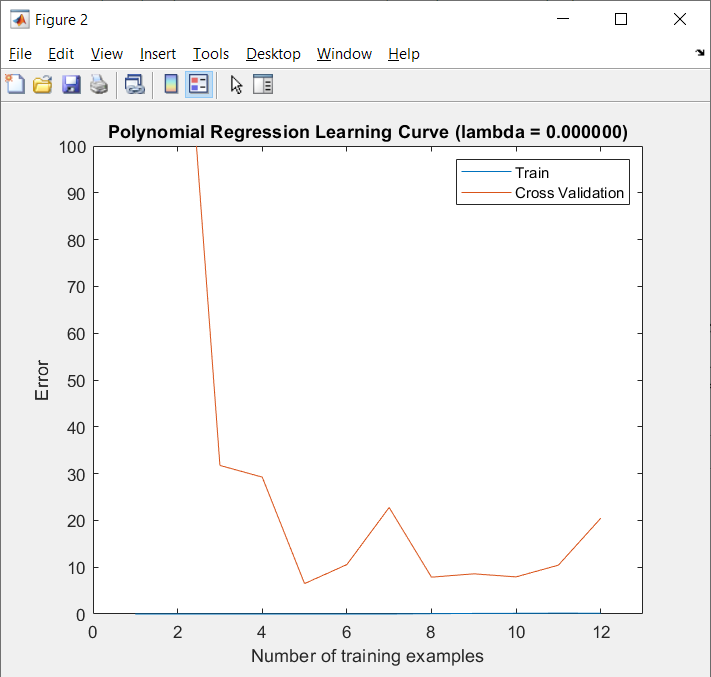


Обучите модель с коэффициентом регуляризации 0 и p = 8.



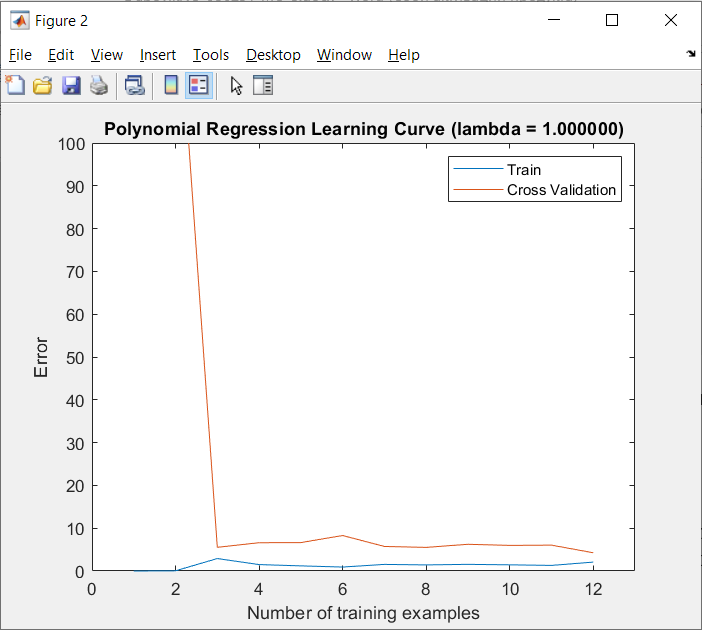
Постройте график модели, совмещенный с обучающей выборкой, а также график процесса обучения. Какой вывод можно сделать в данном случае?

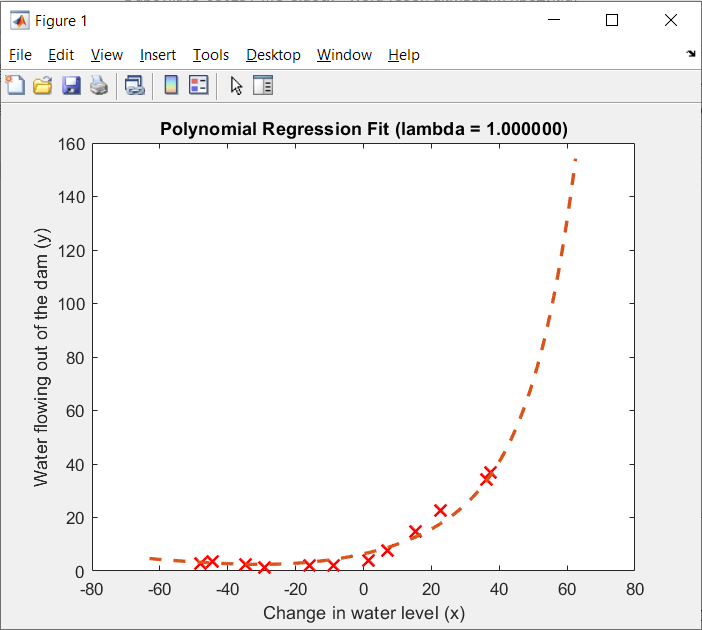


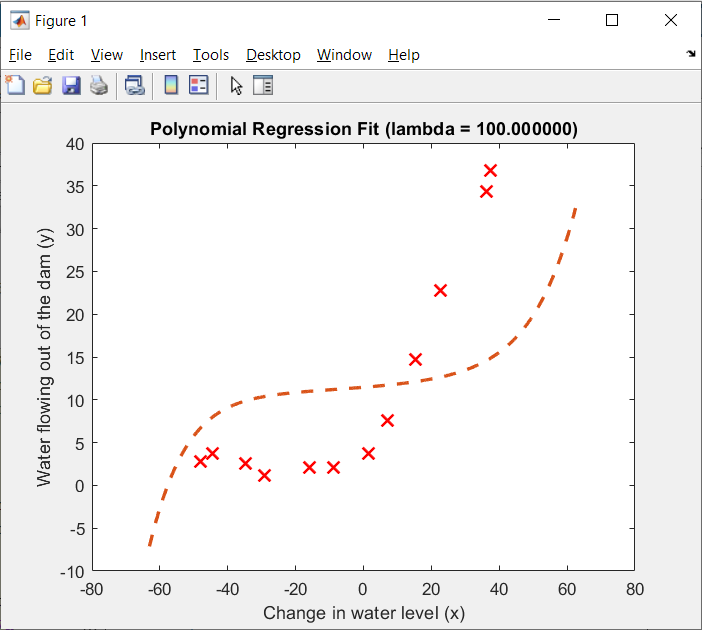


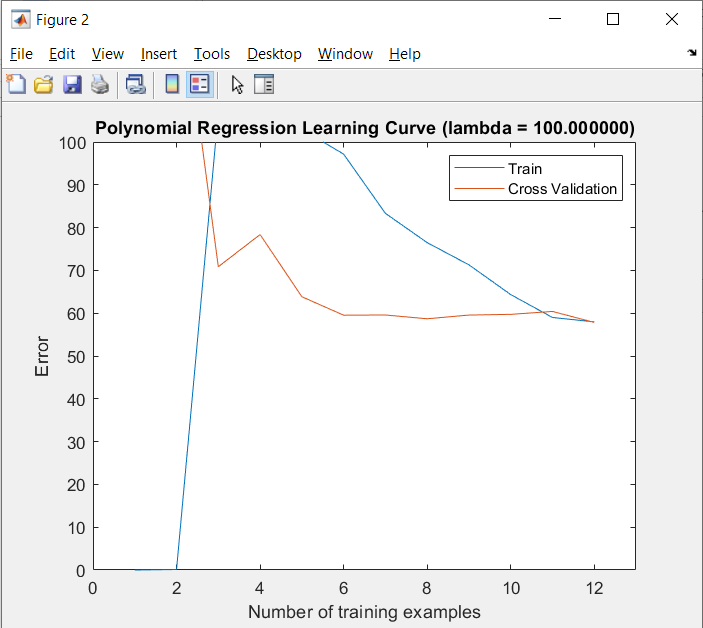
Можно сделать вывод, что полиномиальная регрессия хорошо подстраивается под данные из-за чего ошибка обучения крайне мала. Но из-за этого такая регрессия будет хорошо работать только с тестовым набором данных и будет плохо работать на настоящих данных.

Постройте графики из пункта 10 для моделей с коэффициентами регуляризации 1 и 100. Какие выводы можно сделать?





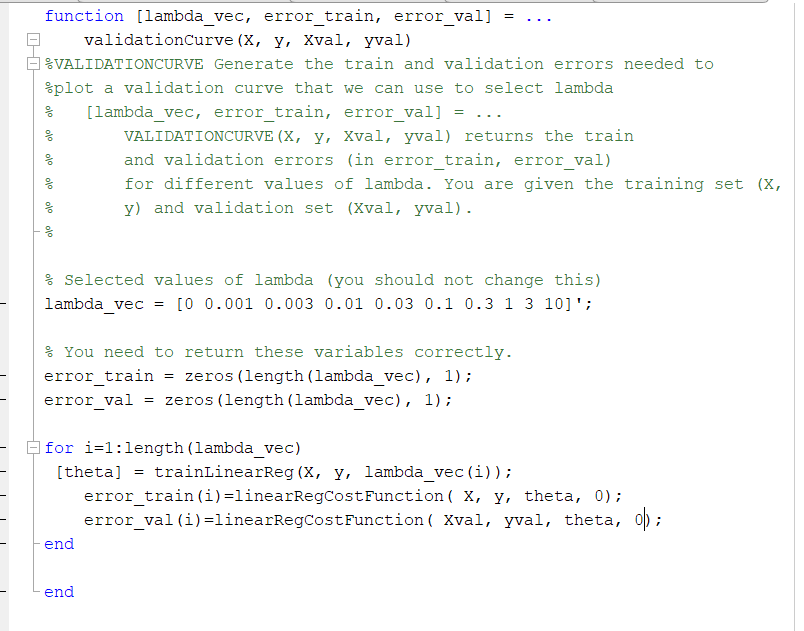


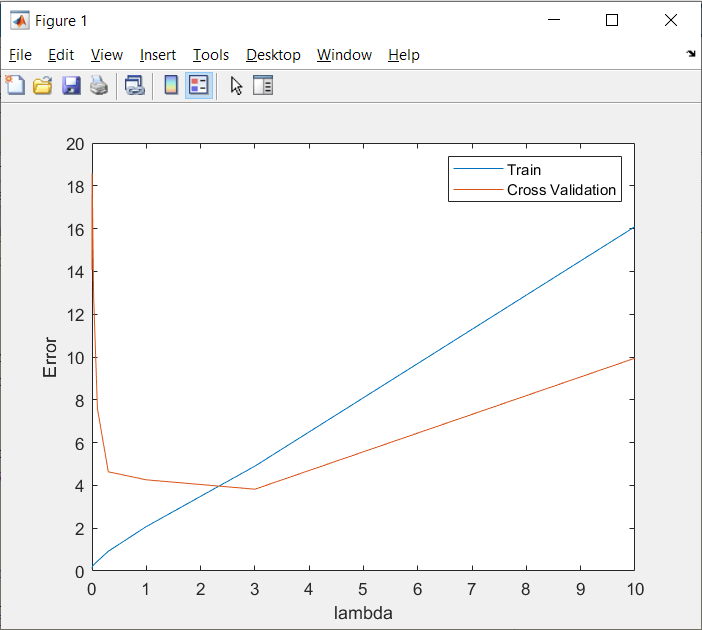


Для λ = 1 можно сделать вывод, что ее полиномиальная модель не имеет больших проблем с отклонением или дисперсией.

В случае λ = 100 полиномиальная линия не следует за данными из-за слишком большой регулизации.

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





Подходящий коэффициент регулизации λ ~ 3

Вычислите ошибку (потерю) на контрольной выборке.

