# ECE521: Assignment 1

Due on Monday, January 25, 2016

**Davi Frossard**

January 25, 2016

# Task 1

| **Training Set Size (N)** | **Validation Errors** |
|---|---|
| 5 | 107 |
| 50 | 40 |
| 100 | 27 |
| 200 | 17 |
| 400 | 9 |
| 800 | 12 |

Using a larger training set increased the algorithm's accuracy since it makes data more diverse and harder to over-fit.

# Task 2

| **Neighbors Considered (K)** | **Validation Errors** |
|---|---|
| 1 | 12 |
| 3 | 8 |
| 5 | 10 |
| 7 | 9 |
| 21 | 11 |
| 101 | 24 |
| 401 | 51 |

Smaller values of K tend to lead to over-fitting when the decision boundary for a variable is complex, that was not the case for this dataset, whose optimum K was 3; on the other hand an over-smoothed decision boundary (larger K) penalized accuracy.

## Task 3

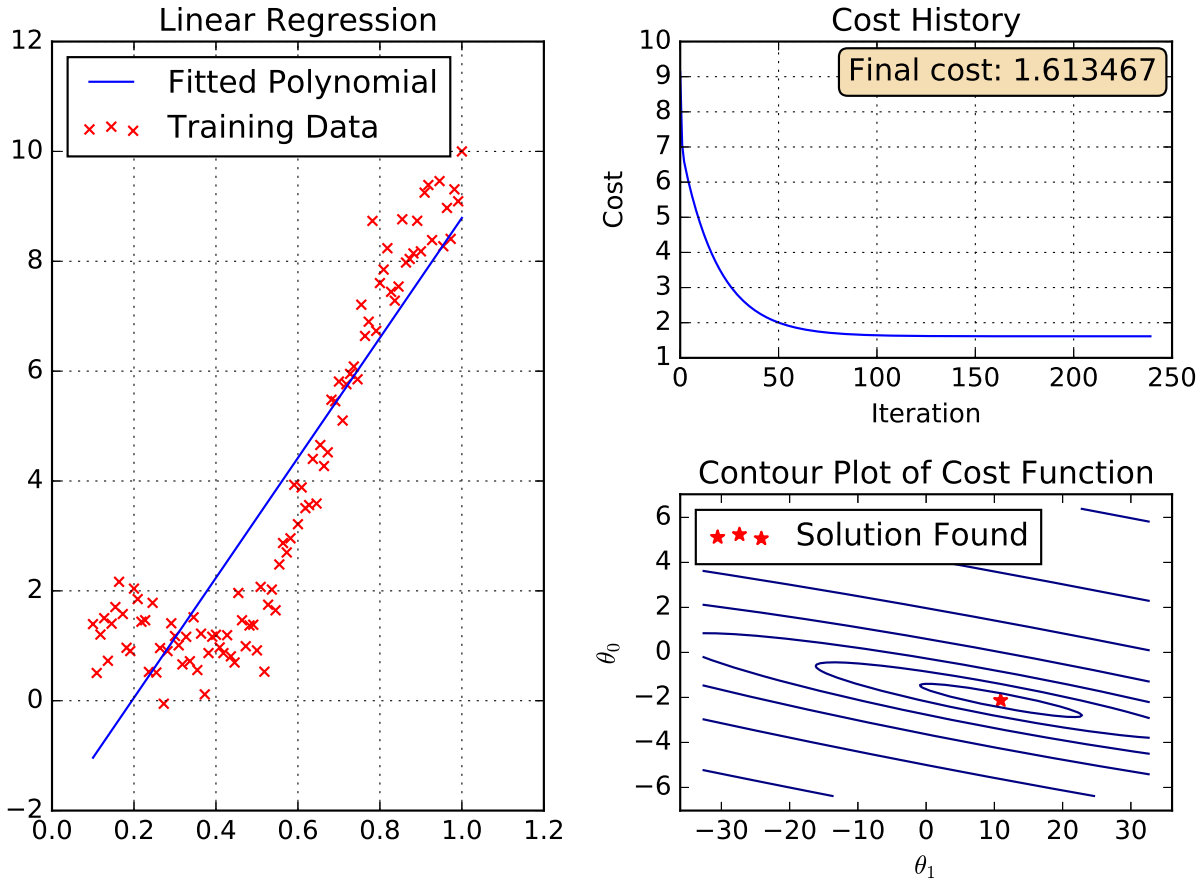


Figure 1: Linear regression of artificial dataset and cost plots

We first normalize the dataset by the maximum value to prevent overflows. Fitting the data to a first order polynomial with gradient descent yields the following model:

$$f(x) = -2.129751 + (10.919595) \cdot x^1$$

## Task 4

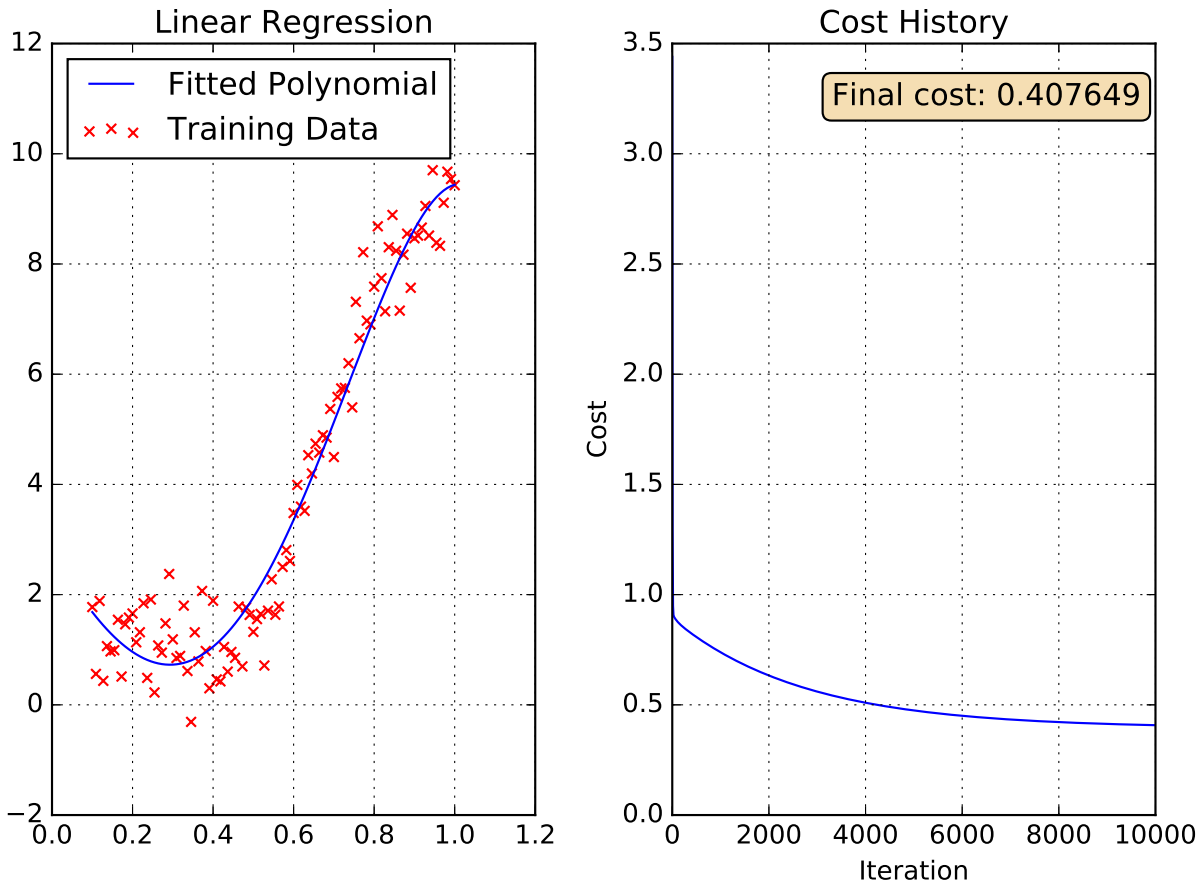


Figure 2: Multivariate regression of artificial dataset and cost plots

By using non linear parameters we give more malleability to the model since it can now make curves, which makes the final cost decrease considerably.

$$f(x) = 2.803670 + (-12.837458) \cdot x^1 + (14.566257) \cdot x^2 + (17.634061) \cdot x^3 + (3.699151) \cdot x^4 + (-16.428233) \cdot x^5$$

# Task 5

| **Training Set Size (N)** | **Validation Errors** |
|---|---|
| 50 | 36 |
| 100 | 30 |
| 200 | 25 |
| 400 | 25 |
| 800 | 24 |

For the same reason as discussed in Task 1, larger values of N provide better results since it provides more samples to build a more accurate model upon and also makes the dataset less prone to over-fitting.

## Task 6

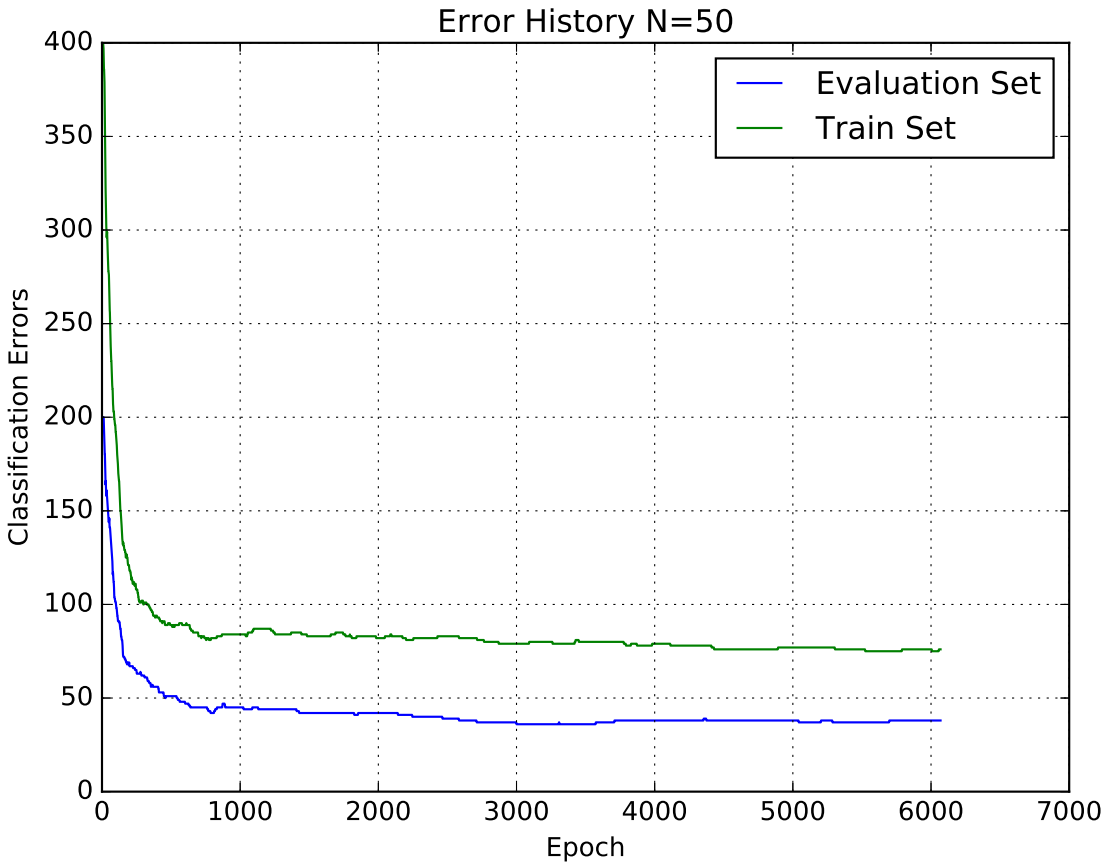


Figure 3: Classification errors by epoch

The behaviour shown in Figure 3 suggests that the model is not suffering from over-fitting.

# Task 7

| $\lambda$ | **Validation Errors** |
|---|---|
| 0.0000 | 36 |
| 0.0001 | 36 |
| 0.0010 | 36 |
| 0.0100 | 37 |
| 0.1000 | 35 |
| 0.5000 | 200 |

Since the model is not suffering from over-fitting (as discussed in Task 6), the smaller values of lambda have no noticeable affect on the accuracy; larger values of lambda, however, lead to under-fitting.