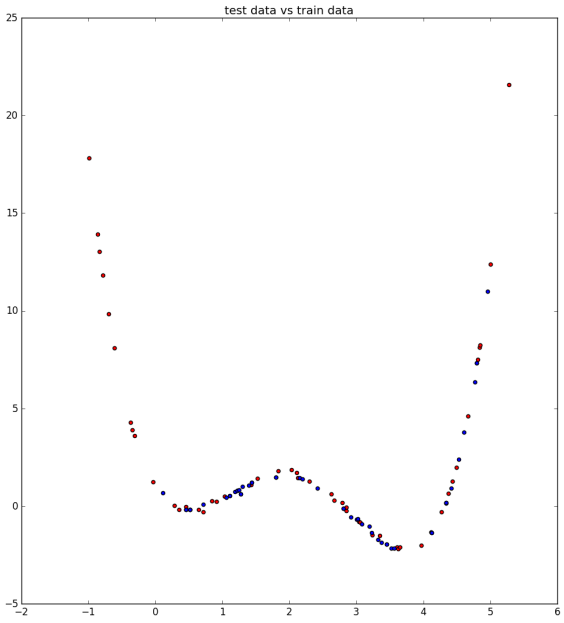
# Homework Report 2

Manuel Del Verme
1769408

December 15, 2016

## 1 Data Preparation

For this experiment a dataset of 50 tulpes (x, y) to be used as training set and a similar one as test set was provided, no regularization was done, centering is done by the zero degree coefficent of the polynomial and x is 1-dimentional.




## 2 Regression

We try to find a law govering the data using a linear fit

## 2.1 Linear regression

A first attempt to fit the data linearly reported an error of: 35.8427, upon visualizing the data it's clear that a polynomial fit might be better suited

## 2.2 Polynomial space transform

### 2.2.1 Choosing a degree

Using a polynomial transform we map our features in a polynomial space while keeping the model linear, the grade of the polynomial has to be choosen taking overfitting in consideration.

| degree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| training ms-error | 35.84 | 25.03 | 27.7 | 23.85 | 23.84 | 23.84 | 23.86 | 23.86 | 23.84 | 23.83 |
| test ms-error | 38.19 | 14 | 95.62 | 0.31 | 0.14 | 0.02 | 1.99 | 5.445 | 157.24 | 4859.53 |

Looking at the error trends, the best polynomial degree would be 4 or 6, the 5th coefficent has a value of 0.0026 hence it's contribution to the regression will be minimal.

### 2.2.2 Best fit

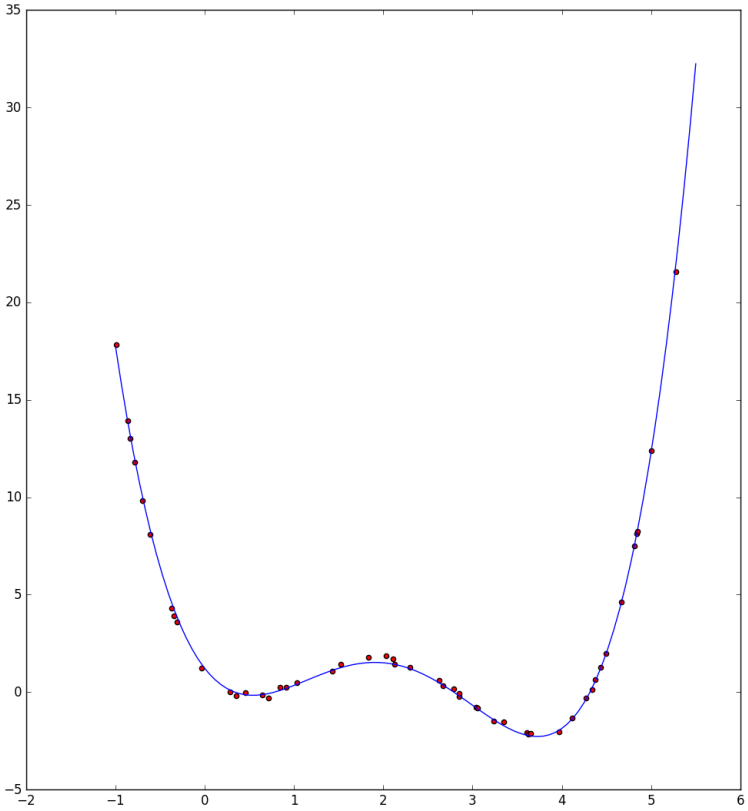



Figure 1: 6th degree fit

### 2.2.3 Training error vs degree

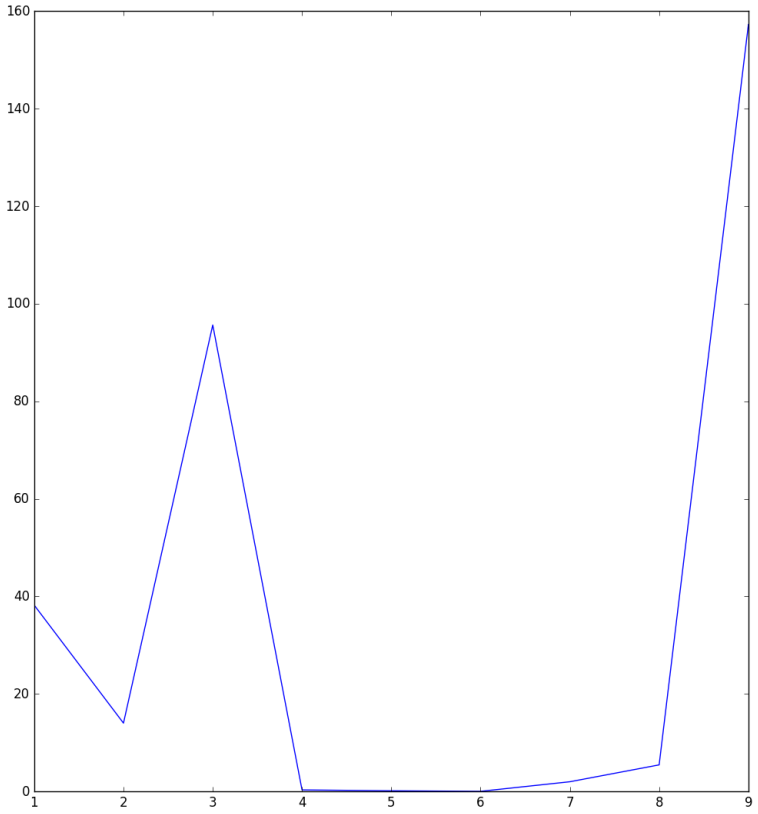



Figure 2: test error