Machine Learning: Second Home Work

Polynomial Regression

Edoardo Ghini

December 14, 2016

Dipartimento di Ingegneria dell'Università di Roma La Sapienza

Contents

I Introduction **1**

1 Scope **1**

2 Objectives **1**

II Development **2**

3 Data loading **2**

4 Training **2**

4.1 Linear Model 2

4.2 Polynomial Model 3

4.3 Error Minimization 3

III Conclusions **4**

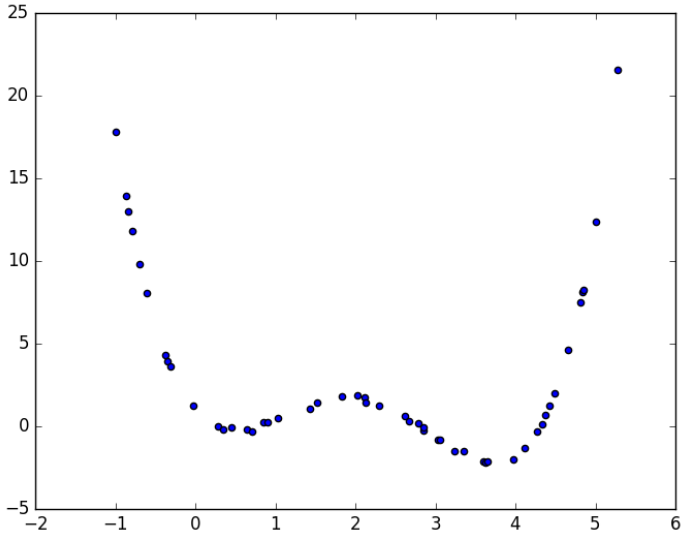


Figure 1

Part I

Introduction

1 Scope

This experience should bring to light the inadequacy of a linear approach when we are treating a non linear distribution of data points.

2 Objectives

It would be required to train a non-linear model with a simple polynomial expansion approach that allows to achieve regression of the dataset in fig(1), chasing a minimum value of the mean square error.

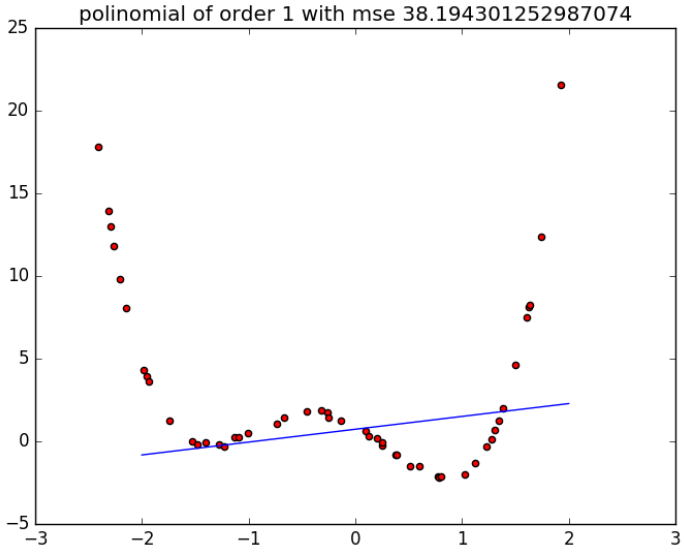


Figure 2

Part II

Development

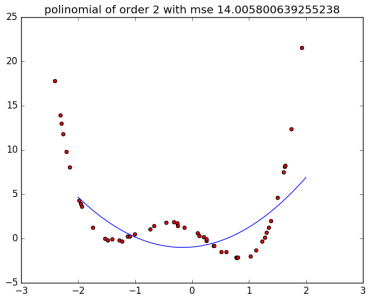
3 Data loading

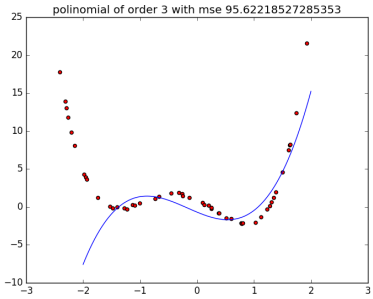
At the beginning , the dataset was loaded, it was already divided in training and test sub-sets and in input values and corresponding labels. Then I standardised the input sub-sets and so, done that, I was ready to approach the training phase.

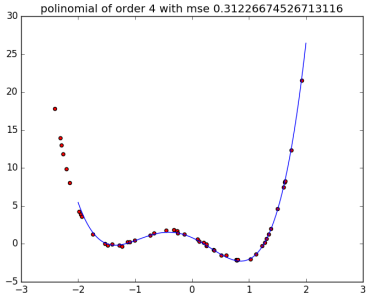
4 Training

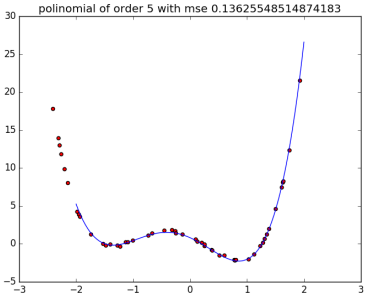
4.1 Linear Model

In order to show the inadequacy of a linear approach I trained a linear regressor with input data and I plotted the result, as shown in fig(2).



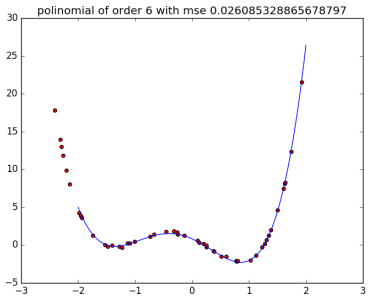


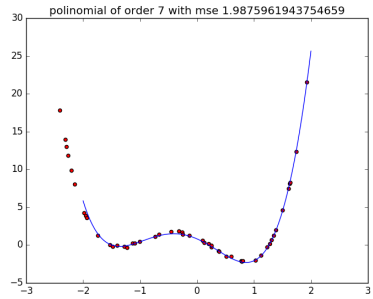


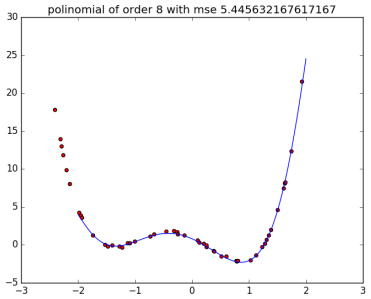


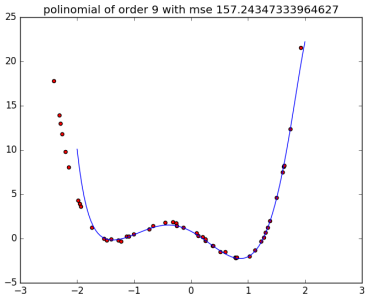
4.2 Polynomial Model

Subsequently I applied a polynomial expansion to input data, that consists in adding terms that are computed by a an exponentiation of the single input terms to obtain a polynomial of higher grade than the beginning. Above, there are some polynomial expansions that I considered in order to find the order that best fits the actual dataset.









4.3 Error Minimization

During the various polynomial grade incrementing iteration I kept track of the mean square error value brought from each polynomial and it turned out that the sixth grade polynomial was the best performer in the regression task. In fig(11) I have represented the mean square error function with the respect of the variation of the grade of the polynomial.

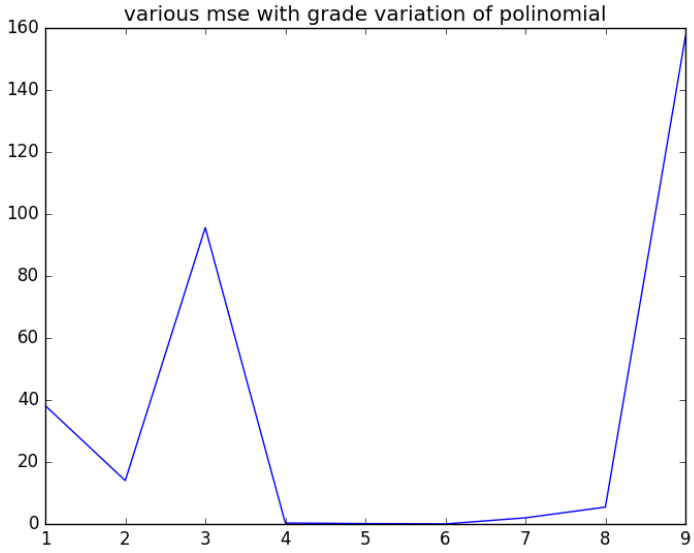


Figure 7

Part III

Conclusions

Finally, considering the collected results, increasing the grade of the polynomial over the threshold, that for this dataset seems to be eight, causes a steep increment of the mean square error. This phenomenon is due to the overfitting of the model indeed. Under those circumstances the model will fit extremely well on the training data but on the other hand it will perform poorly on the test datas.