Machine Learning (CE717-2)

Prof. Soleymani

Assignment #2 - Code Implementation, Due on Aban 1st

Nima Mohammadi

nima.mohammadi@ut.ac.ir

1 Regression without Regularization - Closed Form

For this section three models are implemented, namely, Linear Regression, Polynomial Regression of third degree and Polynomial Regression of fifth degree. The parameters are found using the closed form (normal equation).
Figure 1.1(a) depicts the true targets for the *test* dataset via a blue 3D wireframe. Nearest neighbour interpolation. Table 1.1 lists the training error and test error for each model. The results indicate that the first model had the worst performance (multiple order of magnitude worse than others). The linear regression model has a significant error and could not fit to the data. Its high training error suggest its structural error and inability to model the data. Figure 1.1(b) shows how the predicted values significantly differed the actual data.
One can not see from the plot (Figures 1.1(c) and 1.1(d)), which polynomial regression model is better, but looking at Table 1.1, polynomial regression of 3rd order has slightly lower error reported and can be regarded as the best model that fitted data.

Table 1.1: Result of regression without regularization obtained by the closed form

Closed Form	**Train Error**	**Test Error**
Linear Regression	2.29e+06	6.47e+06
Polynomial Regression (deg=3)	3.31e-20	5.65e-20
Polynomial Regression (deg=5)	1.40e-12	2.91e-12

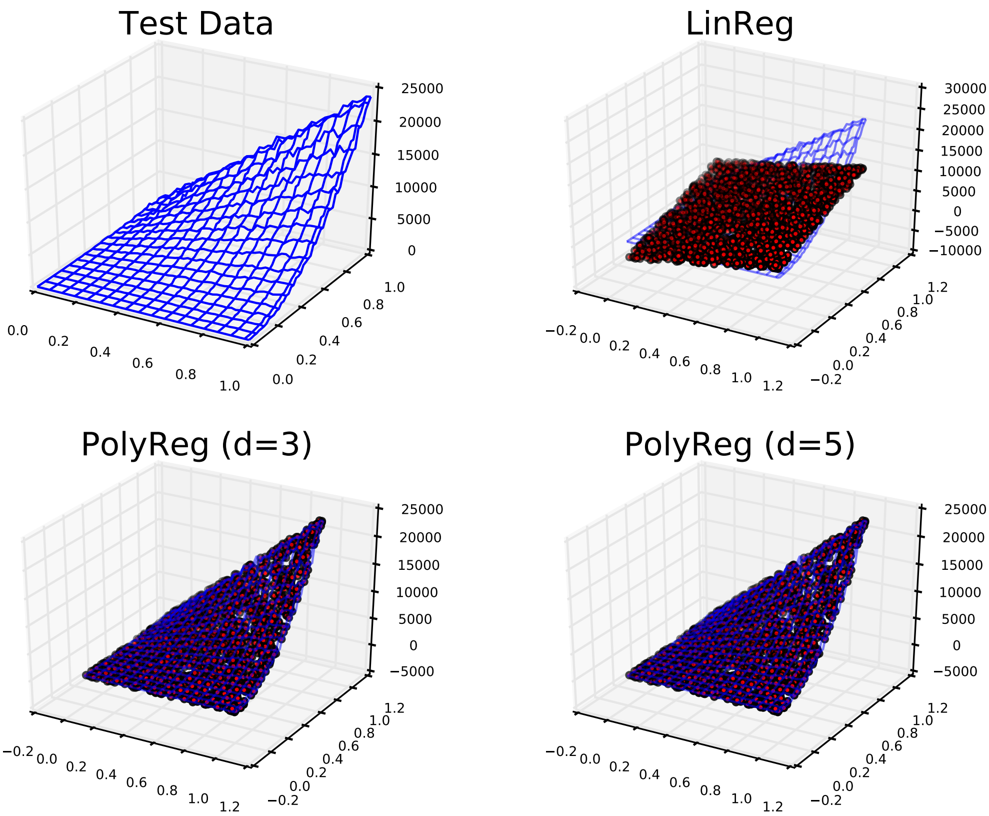

Figure 1.1: (a) Test data (b) Predictions by Linear Regression (c) Predictions by Polynomial Regression of third degree (d) Predictions by Polynomial Regression of fifth degree

2 Regression without Regularization - Gradient Descent

In this section, *Gradient Descent* have been used to find the parameters (i.e. weights) of the model, instead of the *closed form* deployed in previous section.
The stopping criterion for all models is set to max iteration of one million. The learning rate η for the first model is set to $10^{-}6$ and $10^{-}3$ for the two remaining polynomial regressions. Train and test error for each model is shown in Table 2.1. As expected, the results are in accordance with Table 1.1 and still suggests the superiority of Polynomial Regression of third degree to the other.
Figure 2.1 depicts how the error calculated by the cost function behaved for these models. The error axis is in logarithmic scale. PR3 has converged after around 800000 iterations to a very minuscule error. Setting the same learning rate for LR would result in its divergence

Table 2.1: Result of regression without regularization obtained by Gradient Descent

Gradient Descent	**Train Error**	**Test Error**
Linear Regression	2.29e+06	6.47e+06
Polynomial Regression (deg=3)	2.51e-20	7.98e-20
Polynomial Regression (deg=5)	1.33e+01	6.79e+01

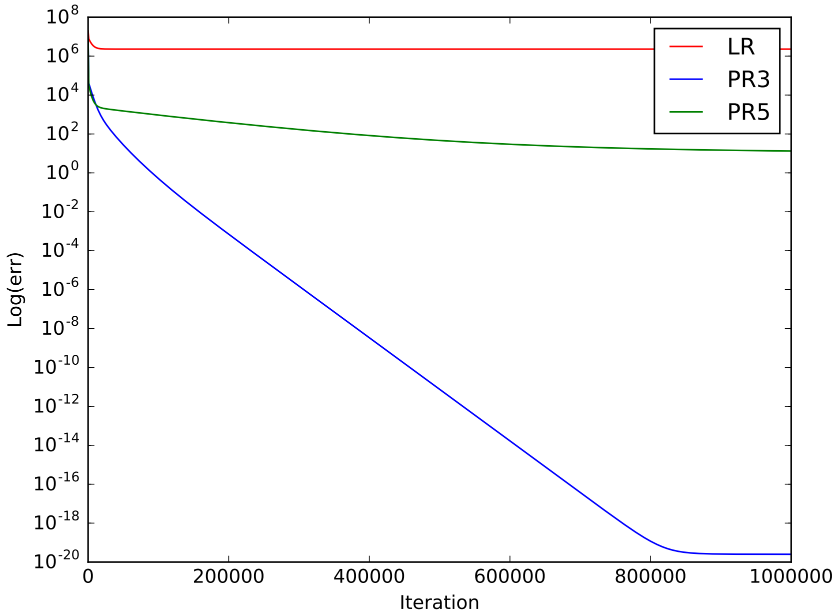

Figure 2.1: The value of cost function during the training over iterations

(hence the lower η). The closed form achieved lower error for PR5, that apparently needs more iterations for this model to be acquired.
It is worth to say that based on empirical results, I have normalized the data beforehand.

3 Regression withRegularization - Closed Form

The best lambda for each models is respectively, 0.1, 0.0001 and 0.0001. Figure 2.2 plots the error value for different regularization parameter λ.
In each plot in Figure 2.2, the average of error for prediction performance on held-out validation folds is drawn, along the train and test errors. The figure indicates that cross-validation is a fine strategy to choose the regularization parameter λ.

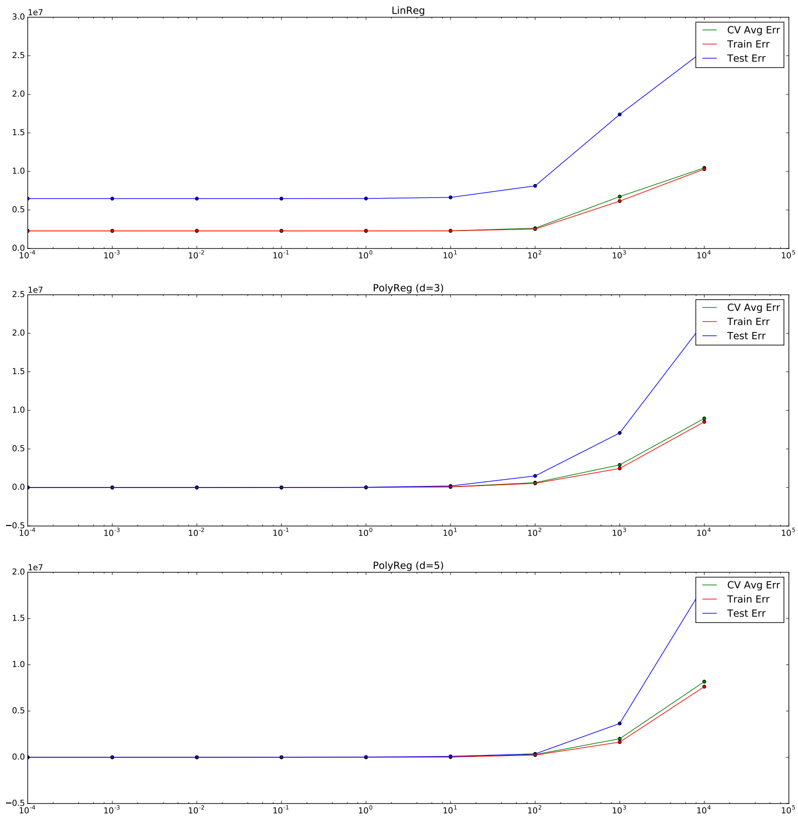

Figure 2.2: Cost for different values of λ