CIS 419/519: Homework 1

{Yongxin Guo}

Although the solutions are entirely my own, I consulted with the following people and sources while working on this homework: Jiangzhu Heng, Yihang Xu

1 Gradient Descent

1. a) Using a constant step size/learning rate alpha can result in potential overshoot around the global minimal, or so-called target value, and jeopardize the performance of convergence
2. b) Using an adaptive step size/learning rate alpha can result in a better convergence when getting closer to the target especially if the learning rate alpha is inversely proportional to the number of iterations.

2 Linear Regression

As noted in the lecture slides:

$$h(x_i) = \theta^T x_i \tag{1}$$

Also,the closed form show that:

$$\theta = (X^T X)^{-1} X^T y \tag{2}$$

Taking the transpose of equation (2) so that we can plug it into equation (1):

$$\theta^T = y^T [(X^T X)^{-1} X^T]^T \tag{3}$$

Then plug it into equation (1), so we get:

$$h(x_i) = y^T [(X^T X)^{-1} X^T]^T x_i \tag{4}$$

which can be re-arranged into:

$$h(x_i) = x_i [(X^T X)^{-1} X^T] y = l_i(x; \mathbf{X}) y_i \tag{5}$$

Thus, the linear regression is a member of this class of estimators

3 Implementing Polynomial Regression

The polynomial regression with $\lambda = 0, \epsilon = 0.001, \alpha = 0.01$ as shown in fig 1.

4 Choosing the Optimal λ

The polynomial regression with $\lambda = 0.0001(tuned), \epsilon = 0.001, \alpha = 0.01$ as shown in fig 2.

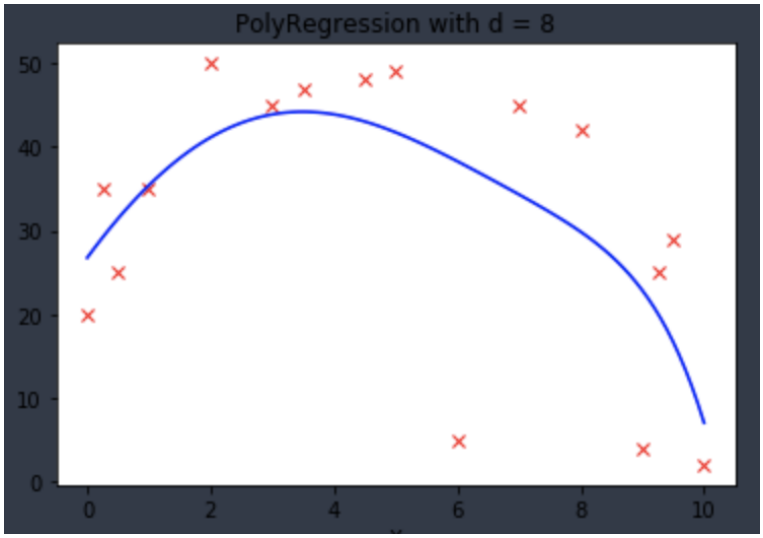

Figure 1: Lambda = 0

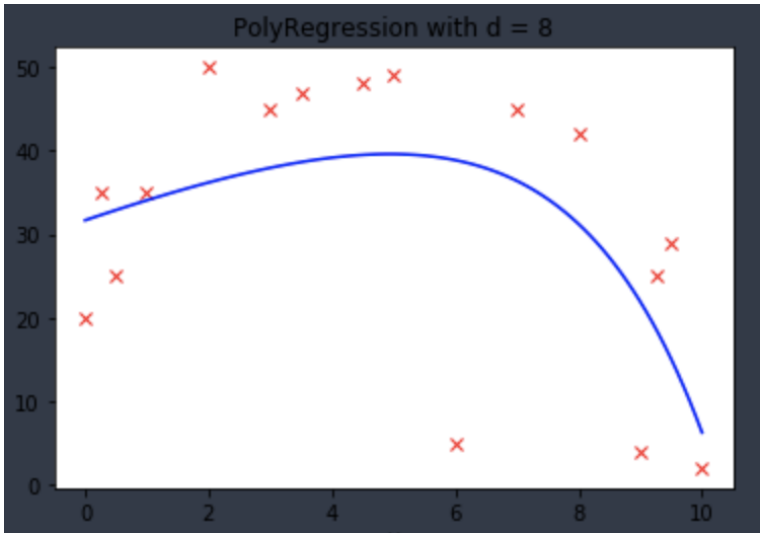

Figure 2: Tuned Lambda