**Assignment 3 :Supervised Learning**

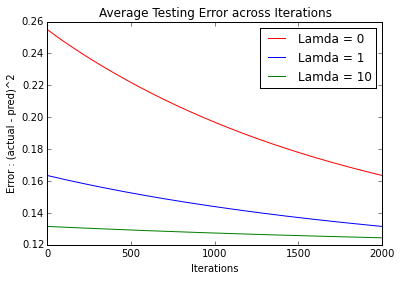
K.Sheetal Reddy

14XJ1A0526

**Deliverables:**

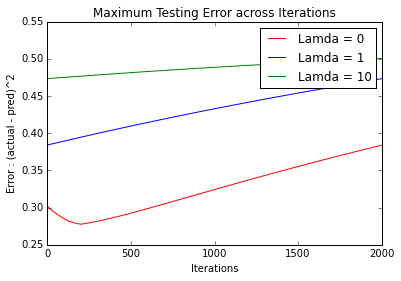
*The first three graphs are for alpha =0.01*

1. **Variations in Average Testing Errors:**



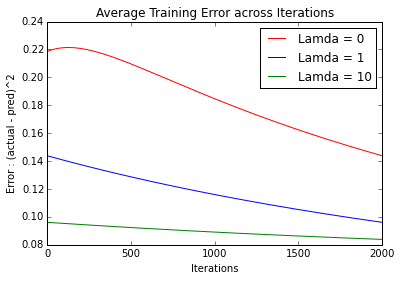
The graph shows how increasing the regularization coefficient, ensures the bounded natures of theta’s making sure that over fitting is prevented. We see the error in case of lambda = 10 reaching a stable horizontal line, which would mean convergence.

1. **Variations in Maximum Testing Errors:**



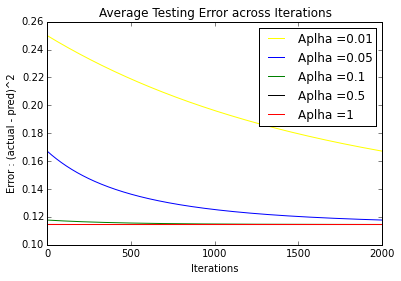
The maximum error is increasing in these cases, because as the thetas improve with iterations, the model predicts the values better. The error begins to reduce and stabilize , but there might be few outliers that the model is unable to fit.

1. **Variations in Average Training Errors:**



As expected the error rates of the training data set are lesser compared to the testing data set. The regularization coefficient pays the same role as discussed in the point (1). We observe a similar trend as the first graph.

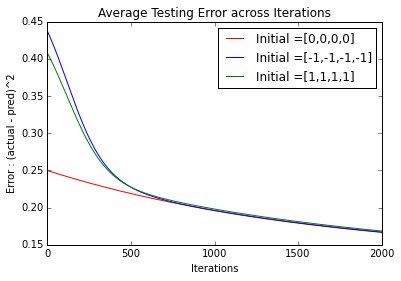
1. **Variations in Learning Rate and Effect on Average Testing Errors**



Alpha is basically the learning coefficient, having a value too high might lead to the solution diverging, the size of the steps taken by the model in the gradient of the slope of the cost function is determined by this parameter. We see how with increasing alpha values the rate at which the value stabilizes(converges) drops drastically.

1. **Variations Initial values and Effect on Average Testing Errors**

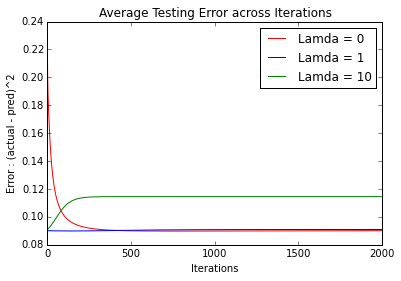
*The first three graphs are for alpha =0.01 and lambda= 10*

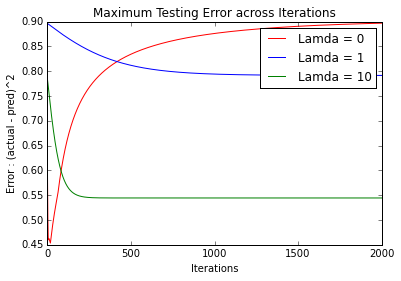


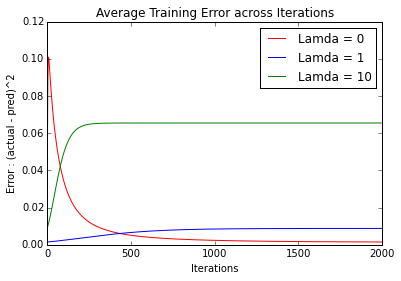
The initial conditions affect the rate at which the model reaches convergence. If we start with values that are closer to the convergence parameter values. The model would require lesser steps and hence lesser iterations to reach the value, than an initial condition from which lies further away from convergence values.

1. **Dynamic Variation of Learning Rate and Effect on the first three graphs:**

The alpha value can also be specifically specified for the different theta values. Hence alpha values can be changed from the usual single value to an array of values for the different parameters. This array can now be modified according to the nature of the slope of the cost function. If the value keeps changing signs it is likely to diverge and the learning rates need to be reduced. If it remains same for long, then it can be increased.







We notice and immediate change in the scale of the values, which is now scattered over a much smaller range than before. We also we can see that the number of iterations required to reach convergence is much lower than before.