SAMYUKTHA RAJ-

PROGRESS REPORT[10/12/2018]-

Completed the first out of the five courses on python(offered by University of Michigan) and finished week one of the second course, and week one of the ML course(by Andrew NG). Relevant codes of the python course one are also attached.

WEEK ONE(LINEAR REGRESSION)-

Regression problems are usually encountered when the target variable that we are trying to predict is continuous. This type of learning algorithm falls under the category of supervised learning.

UNIVARIATE LINEAR REGRESSION-

The aim in this type of algorithm is to find the line of best fit that can be used to predict the output for new un- labelled data. The function that we use to predict(that the algorithm comes up with after training on many labelled data sets) is called the hypothesis function(htheta). The accuracy of the hypothesis function is measured using the cost function. This function calculates the squared error term and the finds the average of all the errors for the given training set.

The number of data points in the given training set is represented by ‘m’. Each data set has an input variable(xi) and an output or target variable(yi). The cost function is given by

*J*(*θ*0​,*θ*1​)=2*m*1​*i*=1∑*m*​(*y*^​*i*​−*yi*​)2=2*m*1​*i*=1∑*m*​(*hθ*​(*xi*​)−*yi*​)2. We move closer to a more accurate hypothesis function by adjusting the values of theta0 and theta1 depending on the value of cost function. When we minimize the cost function, the corresponding values of theta0 and theta1 is considered as the updated values(both have to be updated simultaneously in gradient descent).

GRADIENT DESCENT-

Gradient descent basically helps us fix the appropriate values of theta0 and theta1.

Here theta0 and theta1 maybe referred to as the parameters of the hypothesis function.

Gradient descent algorithm-

Repeat until convergence{

`

}

Where α=learning rate(smaller steps along the curve if alpha is small and larger is alpha is large)

If α is too small the process of minimizing the cost function will take too much time and if its too large the algorithm may never converge. Therefore an optimum alpha must be chosen.It is not necessary to vary alpha as the algorithm progresses as when the slope of the curve becomes less steep(change in slope decreases), the partial differential term automatically decreases thus causing the change in thetaj to be very small and thus converging the algorithm. The derivative also tells us in which direction to proceed i.e if we should increase or decrease the value of theta depending on the relative position wrto global(In case of linear regression)or local minimum of the cost function.

FINDING THE OUTPUT VARIABE USING MATRIX-

Instead of predicting the output variable for each data set individually we can predict the output for the entire training set(in a batch) using matrices.