ML Assignment-2

Deepesh Bhageria(S20170010041)
Rahul Prasad(S20170010118)

Linear regression

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

Perceptron

Linear classifier for binary classification.
A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, belongs to some specific class. It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector.

Gradient descent

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent.

Gradient Descent algorithm :-

*At each iteration j, one should simultaneously update the parameters $\Theta = (\Theta_0\ \Theta_1\ \Theta_2\ \ldots\ldots\ \Theta_n)^t$

*repeat until convergence{

$$\Theta_j = \Theta_j - \eta . \frac{\delta J(\Theta)}{\delta \Theta_j}$$

}

for Linear Regression :-

$$\frac{\delta J(\Theta)}{\delta \Theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h_\Theta(x^{(i)}) - y^{(i)})\, x_j$$

where, m = size of the training set
'j' goes from 0 to n

Data set Description

Group57_regression.txt :- This dataset is for question-2(a).In this dataset we have 308 rows and 7 columns.The 7th column will tell us the 'y' value of that particular row instance.

Group57_perceptron.txt :- This dataset is for question-2(b).In this dataset we have 150 rows and 5 csv values.The 5th column will tell us the class of that particular row instance.

Algorithm

Question - 2(a)

Hypothesis equation = $\Theta^t X$

In this problem we iterate 20000 times,and each time we update the value of theta.
Each time we update the value of theta the hypothesis equation will change.
After 20000 iterations we got the value of theta (by using gradient descent algorithm) which we are using to get the value of test data and after 20000 iterations the gradient becomes very small (< 10^{-5})

Question - 2(b)

Hypothesis equation = W^t X

cost function = $\frac{1}{2} \sum_{j=1}^{m} (t_j - W.X)$

Where,

t_j = Target class
W.X = Hypothesis function

This problem is almost same as linear regression problem.
But in this problem we have taken classes as +1 and -1 for Iris-versicolor and Iris-virginica respectively.
And if hypothesis function(W.X) returns a value greater than equal to zero ,we assume the value to be +1 ,else -1.

Results

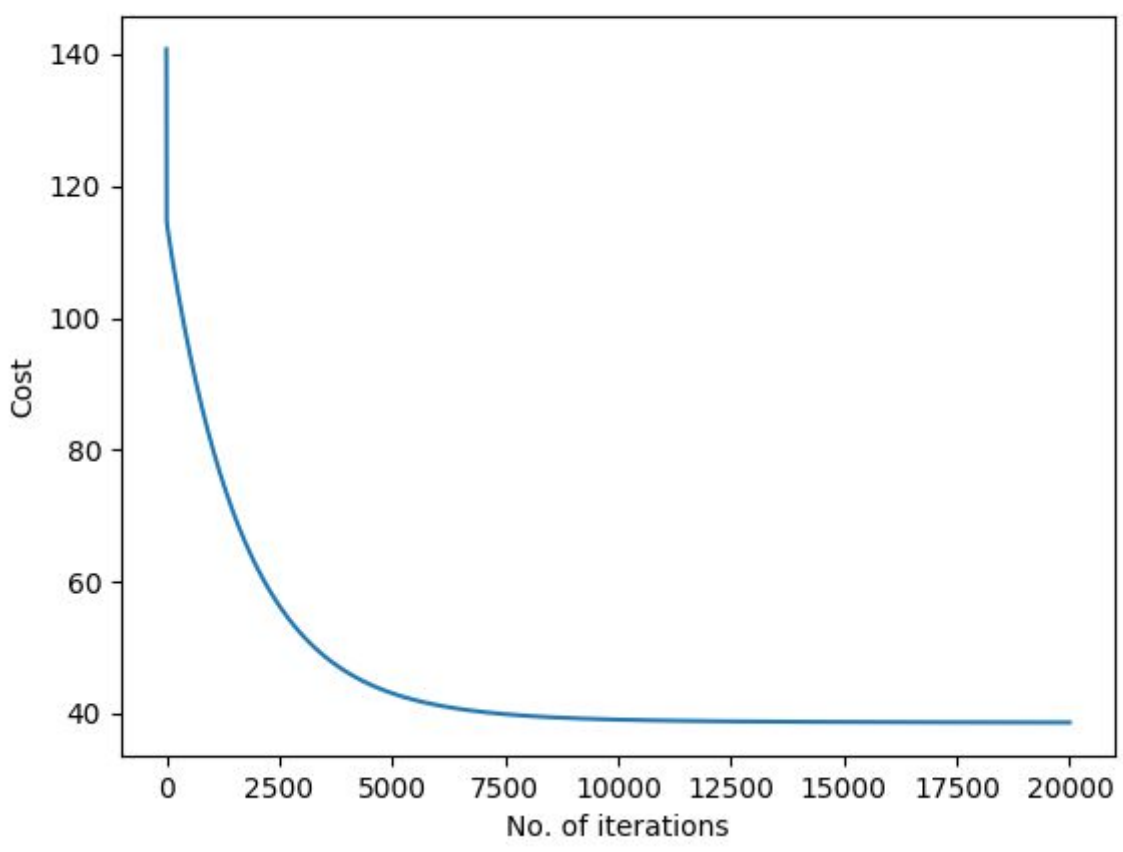
For Question-2(a):

Training error= 38.623351424878024
Testing error 40.540192316148705

Training error= 33.49509343200285
Testing error 51.980160186318734

Training error= 39.003048802451985
Testing error 40.947275121561 13

Training error= 34.7067554236683
Testing error 48.59655121211646



For Question-2(b):

error on training set= 0.0285714285714285 7
error on testing set= 0.0666666666666667

error on training set= 0.1142857142857142 8
error on testing set= 0.0

error on training set= 0.0571428571428571 4
error on testing set= 0.1

error on training set= 0.0571428571428571 4
error on testing set= 0.0

Most of the time we are getting zero error on the test data