Regression - Assignment 1

Answer 1:

a)

Business applications where Machine Learning can be applied are,

I. Predicting CO2 emission for next 5 years grouped by Country

II. Classifying Email as Spam, Grouping information on Web into predefined categories.

III. Predicting weather with >60% probability for a week

Business applications where Machine Learning cannot be applied or yet not proven to be successful are,

I. Language translation with high accuracy (contextual problem)

II. Provide suggestions for controlling rise in temperature due to Global Warming

III. Discover Gold Mine with high accuracy

b)

Unsupervised learning relates input features / parameters to a target pattern and/or groups. Here are the examples,

I. Find market segment for a mid-size SUV in a given area

II. Analyse spending pattern of households in a given region

III. Popularity of a course amongst different age groups.

c)

Classification problems relate the input features to discrete target values. Whereas, a Regression problem relate the input features to continuous value.

They primarily differ in the resultant target variable. In case of Classification problem, the result takes only discrete values such as “True” or “False”, “Yes” or “No”, “Spam” or “Priority” etc.

In case of Regression problem, the target variable takes continuous and infinite values. For e.g. price of a used car.

Examples of Classification Problem,

I. Email spam filter

II. Performance rating system

III. Classifying web information into predefined categories

Examples of Regression problem,

I. House price prediction based on area and location

II. Predict used car price

III. Time to drive to a destination

Answer 2:

The sample data is,

Advt.

Spend Sales

40 920

30 790

25 700

18 580

A simple linear regression expression can be written as, Y = w0 + w1\*X, where w0, w1 are model parameters.

The problem is to find ‘w0’ and ‘w1’ parameter such that it results into minimal Squared Error function.

The squared error function is defined as,

SE = (y1 - (w0+w1x1))2 + (y2 - (w0+w1x2))2 + … + (yn - (w0+w1xn))2

which can be expanded as,

SE(w0, w1) = y12 + (w0+w1x1)2 - 2y1(w0+w1x1) + … +

yn2 + (w0+w1xn)2 - 2yn(w0+w1xn)

SE(w0, w1) = y12 + w02 + (w1x1)2 + 2w0w1x1 - 2w0y1 - 2w1y1x1 +

….

yn2 + w02 + (w1xn)2 + 2w0w1xn - 2w0yn - 2w1ynxn

SE(w0, w1) = (y12 + … + yn2) + nw02 + w12(x12 + x22 + … + xn2) + 2w0w1(x1 + …. + xn) - 2w0 (y1 + …. + yn) - 2w1 (x1y1 + …. + xnyn)

So the definition of squared error equation for a 2 variable linear equation is,

*SE(w0,w1) = n \* mean(y2) + n \* w02 + n \* w12 mean(x2) + 2w0w1 \* n \* mean(x) - 2w0 \* n \* mean(y) - 2w1 \* n \* mean(xy)*

If we substitute the values to derive the Squared error function for the above linear function.

N = 4,

SE(w0,w1) = 2296900 + 4w02 + 3449(w1)2 + 226(w0w1) - 5980(w0) - 176880(w1)

For testing, lets say, w0 = 100, w1 = 20, replacing values, gives us error = 32900, which should match the below calculation.

Error function ED(w) can be derived using least mean square,

ED(w) = 1/2 ∑i=1..N(Yi^ - Yi)2, where our N=4,

Lets assume, w0 = 100, w1 = 20, i.e Y=100+20X

Y^1 = 100 + 20\*40 = 900

Y^2 = 100 + 20\*30 = 700

Y^3 = 100 + 20\*25 = 600

Y^4 = 100 + 20\*18 = 460

ED(w) = 1/2 [(900 - 920)2 + (700 - 790)2 + (600 - 700)2 + (460 - 580)2]= 32900/2

Hence, for the above equation, the Error function is

*SE(w0,w1) = 2296900 + 4w02 + 3449(w1)2 + 226(w0w1) - 5980(w0) - 176880(w1)*

Answer 3:

The equation to solve is,

a11 x1 +a12 x2 +a13 x3 =b1  
a21 x1 +a22 x2 +a23 x3 =b2  
a31 x1 +a32 x2 +a33 x3 =b3

a11 a12 a13 x1 b1

a21 a22 a23 x x2 = b2

a31 a32 a33 x3 b3

And the equation can be solved as,

A x X = B, and X = A-1B

Inverse of a matrix exists only if the Determinant of the matrix > 0. If Determinant of the matrix is <=0, then the equation cannot be solved. And the such equations typically represent parallel lines if we graph them.

Hence to solve simultaneous linear equations, it is mandatory that coefficient matrix has Determinant > 0. When a matrix has |D| > 0, that means, the curves represented by the linear equation meet at some points and hence there exists a solution.

For the Python code to solve simultaneous equation, see the code attachments.