North South University

Department of Electrical and Computer Engineering

Name: Md. Rifat Ahmed

ID: 1931725042

Course: CSE445 – Machine Learning

Section: 6

ASSIGNMENT #1

Submitted To: Dr. B. M. Mainul Hossain (BMH)

Submission Date: 22/3/23

Answer to the Question No - 1

a) This is a Regression problem where the salary of the CEO is the dependent variable which depends on the 3 independent variables profit, number of employees & the industry. Since there's 500 firms in the US so depending on those 3 independent variables around those 500 firms the CEOs salary would be affected.
So, we're not the most interested in predicting the CEOs salary here.

b) This is a Classification problem since we're trying to classify the product into 2 categories success or failure. Also since we're trying to know if the product will be a success or failure, we're most interested in prediction.
There were previously 20 similar products launched and the factors affecting them were price charged, marketing, competition price and 10 other variables. So, a total of 13 variables were deciding whether the product would be a success or failure.

c) This one's another Regression problem. Since we're trying to predict the % change in the USD/Euro exchange rate which can't be classified. Since the exchange rate is being predicted with respect to the weekly change in stock market, so we ~~have~~ are given 1 years worth meaning 52 weeks of data. Now, the number of factors or variable the change is going to depend on is 3 → % change in US market, % change in British market and the % change in German market.

Answer to the Question No - 2

We use Regression analysis to describe the relationships between a dependent variable and one or more independent variables in a system. Specially when we're trying to work with a quantitative output.

It should be avoided when working with non-linear data and when we're expecting discrete outputs because in that case that becomes a classification problem.

Some examples of Regression analysis would be:

* Predicting the effectiveness of marketing or branding of a business. Regression analysis could be used to ~~know~~ predict it the investment in marketing or branding will bring them enough profit purely based on their statistical data.
* It could be used to predict growth plans of a company, predicting their sales volume, predicting inventory status etc.
* So, it could be used to predict how a dependent variable is going to act based

only on the change of the independent variables be it someone's height or age or measuring a factor for a company like marketing or branding.

Answer to the Question No-3

A scatterplot is used to show the relationships between two variables. By the patterns of a scatterplot the relationship type between the variables can be addressed. Such as when a scatterplot shows somewhat of a straight line pattern we can say it has linear relat-ionship but when the scatterplot shows no type of patterns that means the variables have no correlation and when there's any other pattern than a straight line that means it has a non-linear relationship. Ideally we always work with the ~~linear relationship~~ variables that shows linear relationship.

Part - 5

Now, if we had to fit our regression model in one of the four given scatterplots it'd definitely be in Model ⓐ, since model a looks like a straight line pattern that means it has a linear relationship.

As for model ⓑ and ⓒ the scatterplot seems to be scattered around having no specific pattern meaning there's no correlation between the variables there, so, these models are no go.

Then the last model ⓓ seems to have a parabolic shape meaning there's a relation there between the variables but that's non-linear.

So, the most accurate model for prediction would be model ⓐ.

Answer to the Question No - 4

a) Since, there's historical data of children's age, heights are given these datas can be used to train a model that can predict a children's height based on their age.
So, this problem is best addressed using a supervised learning algorithm.

b) We're given the gender and articles of different authors and to predict the gender of a new manuscript's author these datas can be trained on the model.
So, this problem is also best addressed using a supervised learning algorithm.

c) A collection of 1000 essays are given and the task is to group these essays automatically into small groups based on how 'similar' or 'related' they are. Now, there's no data given that can be used to know what makes them similar.
Meaning this problem is best addressed using an unsupervised learning algorithm.

d) To discover if there are sub-types of spam email among a large collection of emails that are known as spam, there's no labelled data available here to divide the spam emails into sub-types.
So, this problem is also best addressed using an unsupervised learning algorithm.

Answer to the Question No - 5

Given,

$$y = \beta_0 + \beta_1 x$$

where, y is the amount of energy released and x is the number of carbon atoms.

We can see from the table that with x increasing the value of y is decreasing that means that the value of β_1 must be negative.

Now, when x is equal to zero, the value of y is equal to β_0 now again if we look at the table we see that for table to follow the decreasing trend the the value of y for $x=0$ must be greater than the value of y when $x=1$. So, β_0 must be greater than -890.

Now, for option (a) β_1 is positive which ~~can't be possible~~ has to be negative for the table to match the linear regression equation, so this option is not possible.

For option (b) the value of β_0 is less than -890 so this option can't be selected either.

Now, if we look at option (c) we can see that the value of β_0 is greater than -890 and the value of β_1 is also negative. So, this could be the right answer.

As for the final option (d), again the value of β_0 is less than -890, so this option is invalid too.

∴ Option (c) has to be the correct answer for the linear regression equation to be used on the given table as the other options does not meet the criteria set for β_0 and β_1.

$$\therefore \beta_0 = -569.6 \quad \& \quad \beta_1 = -530.9$$