26 March Assignment

June 15, 2023

[ ]:
```
Q1. Explain the difference between simple linear regression and multiple linear␣
 ↪regression. Provide an
example of each.
```

ANS-

[ ]:
```
Simple Linear Regression :-

Simple Linear Regression establishes the relationship between two variables␣
 ↪using a straight line. It attempts to draw a line that comes
closest to the data by finding the slope and intercept which define the line␣
 ↪and minimize regression errors. Simple linear regression has
only one x and one y variable.

Multi Linear Regression :-

Multiple Linear regressions are based on the assumption that there is a linear␣
 ↪relationship between both the dependent and independent
variables or Predictor variable and Target variable. It also assumes that there␣
 ↪is no major correlation between the independent variables.
Multi Linear regressions can be linear and nonlinear. It has one y and two or␣
 ↪more x variables or one dependent variable and two or more
independent variables.
```

[ ]:
```

```

[ ]:
```
Q2. Discuss the assumptions of linear regression. How can you check whether␣
 ↪these assumptions hold in
a given dataset?
```

ANS-

[ ]:
```
The assumptions of linear regression are:

1.Linearity: The relationship between the independent and dependent variables␣
 ↪is linear.
2.Independence: The observations are independent of each other.
```

```
3.Homoscedasticity: The variance of the residuals is constant across all levels␣
 ↪of the independent variable.
4.Normality: The residuals are normally distributed.

To check whether these assumptions hold in a given dataset, you can use the␣
 ↪following methods:

1.Linearity: You can check for linearity by evaluating a Residuals vs Fitted␣
 ↪plot.
2.Independence: Each observation should be independent of one another.
3.Homoscedasticity: You can check for homoscedasticity by evaluating a␣
 ↪Residuals vs Fitted plot.
4.Normality: You can check for normality by evaluating a Normal Q-Q plot.
```

[ ]:

[ ]:
```
Q3. How do you interpret the slope and intercept in a linear regression model?␣
 ↪Provide an example using
a real-world scenario.
```

ANS-

[ ]:
```
In a linear regression model, the slope represents the rate of change of the␣
 ↪dependent variable with respect to the independent variable.
The intercept represents the value of the dependent variable when the␣
 ↪independent variable is zero. For example, if you want to model how
the cost of renting a car depends on the number of days, you can use a linear␣
 ↪equation with slope and intercept. The slope represents the
rate of change of the cost per day, and the intercept represents the fixed fee␣
 ↪or deposit.

Here's an example: Suppose you are a real estate agent and you want to predict␣
 ↪the price of a house based on its size. You collect data on
house prices and sizes in your area and fit a linear regression model to this␣
 ↪data. The slope of this model represents how much the price of
a house increases for each additional square foot of living space. The␣
 ↪intercept represents the price of a house with zero square feet of
living space, which is not meaningful in this context.
```

[ ]:

[ ]:
```
Q4. Explain the concept of gradient descent. How is it used in machine learning?
```

ANS-

[ ]:
```
Gradient descent is an optimization algorithm that is used to minimize the loss␣
 ↪function or the cost function in a machine learning model.
```

```
The algorithm works by iteratively adjusting the parameters of the model in the␣
 ↪direction of the steepest decrease in the function.
The steepest decrease is determined by calculating the gradient of the function␣
 ↪with respect to the parameters. The gradient is a vector
that points in the direction of the greatest increase in the function. By␣
 ↪taking the negative of this vector, we can move in the direction
of the greatest decrease in the function. This process is repeated until a␣
 ↪minimum of the function is reached.

In machine learning, gradient descent is commonly used to train models and␣
 ↪neural networks. Training data helps these models learn over time,
and the cost function within gradient descent specifically acts as a barometer,␣
 ↪gauging its accuracy with each iteration of parameter updates.
```

[ ]:

[ ]:
```
Q5. Describe the multiple linear regression model. How does it differ from␣
 ↪simple linear regression?
```

ANS-

[ ]:
```
Simple linear regression is a statistical method that uses a single feature to␣
 ↪model a linear relationship with a target variable.
Multiple linear regression is a more specific calculation than simple linear␣
 ↪regression. It uses multiple features to model a linear
relationship with a target variable. Multiple linear regression is often better␣
 ↪for more complex relationships requiring more consideration.

In simple linear regression, there is only one independent variable impacting␣
 ↪the slope of the relationship. In contrast, multiple regression
incorporates multiple independent variables. Each independent variable in␣
 ↪multiple regression has its own slope coefficient.
```

[ ]:

[ ]:
```
Q6. Explain the concept of multicollinearity in multiple linear regression. How␣
 ↪can you detect and
address this issue?
```

ANS-

[ ]:
```
Multicollinearity is a phenomenon in which two or more independent variables in␣
 ↪a multiple regression model are highly linearly related.
It can cause problems when you fit the model and interpret the results.

The best way to detect collinearity in the linear regression model is the␣
 ↪multicollinearity variance inflation factor (VIF), calculated to
```

```
figure out the standard of tolerance and assess the degree of collinearity. For␣
 ↪example, if the VIF is 4, indicating a tolerance of 0.25 or
lower, there is a possibility that the phenomenon will occur.

There are several ways to address multicollinearity in multiple linear␣
 ↪regression. Some of them are:

1.Remove one of the correlated variables.
2.Combine the correlated variables into one variable.
3.Use principal component analysis (PCA) to transform the correlated variables␣
 ↪into uncorrelated principal components.
```

[ ]:

[ ]:
```
Q7. Describe the polynomial regression model. How is it different from linear␣
 ↪regression?
```

ANS-

[ ]:
```
Polynomial regression is a type of regression analysis that is used when the␣
 ↪relationship between the independent variable x and dependent
variable y is not linear. In polynomial regression, the relationship between␣
 ↪the independent variable x and dependent variable y is modeled
as an nth degree polynomial in x. Linear regression, on the other hand, is used␣
 ↪when the relationship between the independent variable x and
dependent variable y is linear. In linear regression, the relationship between␣
 ↪x and y is modeled as a straight line.

In other words, polynomial regression is used when the relationship between x␣
 ↪and y cannot be accurately modeled by a straight line.
Polynomial regression can be used to model relationships that are curved or␣
 ↪have more complex shapes than a straight line.
```

[ ]:

[ ]:
```
Q8. What are the advantages and disadvantages of polynomial regression compared␣
 ↪to linear
regression? In what situations would you prefer to use polynomial regression?
```

ANS-

[ ]:
```
Polynomial regression is a special case of multiple linear regression. The␣
 ↪relationship between the independent variable x and dependent
variable y is modeled as an nth degree polynomial in x. Linear regression␣
 ↪cannot be used to fit non-linear data (underfitting).
Therefore, we increase the model's complexity and use Polynomial regression,␣
 ↪which fits such data better.
```

```
The advantages of polynomial regression are that it can fit a wide range of␣
 ↪functions and can be used to model non-linear relationships between
variables. It can also be used to model interactions between variables.

The disadvantages of polynomial regression are that it can be sensitive to␣
 ↪outliers and can overfit the data if the degree of the polynomial
is too high.

Polynomial regression is preferred over linear regression when the relationship␣
 ↪between the independent variable x and dependent variable y
is non-linear. It is also preferred when there are interactions between␣
 ↪variables.
```