1. In a linear equation, what is the difference between a dependent variable and an independent variable?

Ans: **the dependent variable is the one which we predict from the Machine Learning Model and independent variable is the one which decides the variation of the dependent variable.**

1. What is the concept of simple linear regression? Give a specific example.

Ans: **the simple linear regression has only 2 variables. One is dependent and the other one is independent. We find the best fit line in linear regression by minimising the mean squared error(MSE) or Mean Absolute Error(MAE). For example we have two variables sales of juice vs timing then here timing is the independent variable and sale of juice is a dependent variable which we predict from the model. Here the linear regression algorithm draws a linear line showing the relationship between these two variables.**

1. In a linear regression, define the slope.

Ans: **in a linear regression equation y=mx+c, slope is denoted by m which basically shows how much the value of y would increase if we increase the value of x by 1.**

1. Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).

Ans: **the slope is .75**

1. In linear regression, what are the conditions for a positive slope?

Ans: **condition is if one variable increases the other one must increase to be able to produce a positive slope.**

1. In linear regression, what are the conditions for a negative slope?

Ans: **upon increasing one variable, the other variable must decrease.**

1. What is multiple linear regression and how does it work?

Ans:  **in multiple linear regression, unlike simple linear regression, we have more than one independent variable and one dependent variable. The equation changes for the multiple linear regression depending upon the number of independent variables. We try to find the best fit line and the equation for the same to predict the values for the target variable. As an evaluation matrix, we use R square and Adjusted R square to see how many variables are able to explain the variation in the target variables.**

1. In multiple linear regression, define the sum of squares due to error.

Ans: **SSE or sum of square is used to identify error in a regression model. Where error is determined by the calculating the distance between all the observed values and predicted values and squaring them up to omit the negative values. The higher the SSE is the more error a model has. Therefore, lower value of SSE is better to be able to build a good regression model. Overfitting in the model will result in SSE being 0 as there will not be any difference between observed values and predicted values but this model won’t work well with unseen values.**

1. In multiple linear regression, define the sum of squares due to regression.

Ans: **SSE or sum of square is used to identify error in a regression model. Where error is determined by the calculating the distance between all the observed values and predicted values and squaring them up to omit the negative values. The higher the SSE is the more error a model has. Therefore, lower value of SSE is better to be able to build a good regression model. Additionally, Overfitting in the model will result in SSE being 0 as there will not be any difference between observed values and predicted values but this model won’t work well with unseen values.**

1. In a regression equation, what is multicollinearity?

Ans: **multicollinearity when there are 2 highly correlated independent features in the dataset. it can cause an issue in ML model as it becomes difficult to find the effect of an individual independent variable on the target variable. To fix this issue, one of the highly correlated independent variable can be omitted from a dataset.**

1. What is heteroskedasticity, and what does it mean?

Ans: **it means the data points are not equally distributed and that is usually occurred due to the presence of extreme values of outliers. That is why one of the assumption of using the linear regression model is homskedasticity.**

1. Describe the concept of ridge regression.

Ans: **Ridge regression model is used when we get overfitted model in linear regression. In the overfitted model, the value of cost function becomes 0 as the distance between the observed and predicted values is zero. So in order to eliminate overfitting we add a hyperparameter to the cost function so the algorithm keeps training despite the value of cost function being zero.**

**The reason why the overfitted mode is undesirable is due to its high error rate with the testing data set. Overfitted model performs well with the training dataset but doesn’t perform well with the testing data set as it tried to remember the value of the training data set.**

**So to conclude, the ridge regression is used when the model is overfitted in linear model.**

1. Describe the concept of lasso regression.

Ans: this regression algorithm is used for feature selection as it identifies which features are not adequately correlated to the target variable. The equation for the model is the below:

Diagram

Description automatically generated

Here the relation between alpha and m is such that as the value of alpha increases, the value of slope decreases and reaches zero. So we consider only those feature those coefficient values are not zero and that is how we do feature selection through lasso regression.

1. What is polynomial regression and how does it work?

Ans: **Polynomial regression has nth degrees or we can say the linear equation is 1 degree polynomial equation. It is mainly used when the relationship between independent variable and dependent variable is non-linear. If the relationship is linear then linear or multiple linear regression can be used depending upon the number of independent variables. The equation of the polynomial regression with the nth degrees is below:**

**y= b0+b1x + b2x2+ b3x3+....+ bnxn**

1. Describe the basis function.

Ans: **while using linear regression, we assume that the relationship between variables is linear. However, if the relationship is non-linear then we use basis functions. For instance, Polynomial regression is used when the relationship between variables is non-linear. The equation looks like below:**

Diagram

Description automatically generated with medium confidence

1. Describe how logistic regression works.

Ans: **logistic regression is used for binary classifications problems. For example, we have a data set of customers who have bought several products online. Now a model needs to be developed to predict the if the user would be satisfied or unsatisfied. Since the output variable is a binary classified variable. Therefore, logistic regression algorithm can be employed.**