#1 What is Linear regression?

'''Linear Regression is a supervised machine learning technique use to predict
values of Target columns by finding linear relationship between dependant and independant

#2 How do you represent a simple linear regression?

'''Simple linear regression has one independant variable and one dependant variable,
equation is y=mx+C where m is the slope associated with independant variable and c is an i

#3 What is multiple linear regression?

'''Multiple linear regression has more than one independant variable and one dependant var
equation is y=m1x1+m2x3+....+mnxn+C where m is the slope associated with independant varia

#4 What are the assumptions made in the Linear regression model?

Assupmtions are as follows

'''
1. There has to be a linear relation of each independant variable with dependant variable

2. There shouldn't be a multicolinearity between independant varibales

3. Normality of residual

4. Homoscedascticity
'''

#5 What if these assumptions get violated?

'''Our Model might get overfitted or underfitted'''

#6 What is the assumption of homoscedasticity?

'''Homoscedasticity means the variance for error term for all independent variable should

#7 What is the assumption of normality?

'''Normality of residual means the mean value of the residual should be nearly equal to 0'

#8 How to prevent heteroscedasticity?

'''Detect and Handle outliers, because this algorithm is outlier sensetive

#9 What does multicollinearity mean?

'''Multicolinearity meaning is one of the independent variable is having linear relation w
if x1 is increasing x2 also increasing'''

#10 What are feature selection and feature scaling?

'''Feature selection is nothing but selecting x and y for model building and feature
scaling means applying Normalization or Standardization on features to scale down them in

#11 How to find the best fit line in a linear regression model?

'''BFL is nothing but which has low bias and variance and High training and
testing accuracy. Passes through maximum data points and have optimum c nad m values'''

#12 Why do we square the error instead of using modulus?

'''If we use absolute value then that equation is not differentiable when we calculate opt
m and c using gradient descent, but if we use squared terms then we can differentiate that

#13 What are techniques adopted to find the slope and the intercept of the linear regressi

'''Gradient descent is use to find the optimum values of m and c,
it differenctiate cost function i.e MSE with respect to m and c and calculate it. In that

#14 What is cost Function in Linear Regression?

'''MSE Mean Squarred Error is the Cost Function of Linear Regression, Sum(Ya-Yp)^2/n'''

#15 briefly explain gradient descent algorithm

'''
Gradient descent is use to find the optimum values of m and c, heance it Gives Best Fit Li
It differenctiate costs function i.e MSE with respect to m and c and calculate best values
In that alpha is used as Learning rate and Learning step will be each step and that will d
Mnew = Mold - alpha*derivation(MSE) w.r.t 'm'
Cnew = Cold - alpha*derivation(MSE) w.r.t 'c'
This is how it will Keep on finding m and c values until
iterations = 10000
epsilon (sqrt.(Mnew^2-Mold^2))<0.001
either of these conditions are getting satisfied
'''

#16 How to evaluate regression models?

'''We generally evaluate regression model based on following metrics:
1. r2_score
2. MSE
3. RMSE
4. MAE
5. model score

these values we calculate for both Train and Test dataset and see if our model is optimal

```
#17 Which evaluation technique should you prefer to use for data having a lot of outliers

'''Mean Absolute Error(MAE) is preferable to use for data having too many outliers in it b
whereas MSE and RMSE are very susceptible to outliers and starts penalizing the outliers b


#18 What is residual? How is it computed?

'''Residual is also called as Error Term which is difference between actual and predicted


#19 What are SSE, SSR, and SST? and What is the relationship between them?

'''
SSE = Sum(Ya-Yp)^2/n

SSR = Sum(Yp-Ymean)^2/n

SST = SSE = Sum(Ya-Ymean)^2/n

Coeff of determination,r_2 score :  1 - (SSE/SST)

when Ya = Yp, r_2 score is 1 Good model

when Ymean = Yp, r_2 score is 0 Bad Model
'''



#20 What’s the intuition behind R-Squared?

'''
Coeff of determination,r_2 score :  1 - (SSE/SST)

when Ya = Yp, r_2 score is 1 Good model

when Ymean = Yp, r_2 score is 0 Bad Model

The r_2 value has one drawback, the value gets increse even if we add bad feature in our d
'''


#21 What does the coefficient of determination explain?

'''Coefficient of determination explains the goodness of bestfit line'''


#22 Can R2 be negative?

'''Note that it is possible to get a negative R-square for equations that do not contain a
Because R-square is defined as the proportion of variance explained by the fit,
if the fit is actually worse than just fitting a horizontal line then R-square is negative


#23 What are the flaws in R-squared?
```

```
'''The r_2 value has one drawback, the value gets increse even if we add bad feature in ou


#24 What is adjusted R2?

'''Adjusted r2 is always less that r2 score and formula is 1- (1-r2)(N-1) / (N-P-1) .
 It gives less value if we are adding any bad feature in our dataset and it gives higher v


#25 What is the Coefficient of Correlation: Definition, Formula

'''
Coefficient of Correlation, R: It shows how each variable are correlated with each other.
and 1 is for ideal positive correlation(max) and -1 for negative correlation(max).

If correlation is greater than 0.7 then it said to be a very good positively correlated va
then they also called as very good neagatively correlated variables.

Formula is, Variance(x,y)/std(x)std(y)

sum.(Xi - Xmean)(Yi - Ymean)/sum.sqrt.((Xi - Xmean)^2.(Yi - Ymean)^2)
'''


#26 What is the relationship between R-Squared and Adjusted R-Squared?

'''R-Squared value is always greater than or equal to Adjusted R-Squared'''


#27 What is the difference between overfitting and underfitting?

'''
Overfitting: When the Training Accuracy is more and Testing Accuracy is less then model is
Model and Bias is also low

Underfitting: When the Training and Testing both accuracies are low that time Variance is
'''


#28 How to identify if the model is overfitting or underfitting?

'''
When the Training Accuracy is more and Testing Accuracy is less then model is overfitted a
is less and Testing Accuracy is less then model is underfitted
'''


#29 How to interpret a Q-Q plot in a Linear regression model?

'''If all the points are lying very near to the linear line approximately on the line then



#30 What are the advantages and disadvantages of Linear Regression?
```

```
'''
Advantages:
1. Easy to implement
2. Easy to understand
3. less complexity to compared to other algorithms.
4. Linear Regression is susceptible to over-fitting but it can be avoided using
   some dimensionality reduction techniques, regularization (L1 and L2) techniques and cro

Disadvantages:
1. Sensitive to outliers
2. Parametric, comply with all the assumption
3. Need to handle null values
4. linear regression is not a complete description of relationships among variables.

'''


#31 What is the use of regularisation? Explain L1 and L2 regularisations.

'''
Regularization is use when we get overfitted model for Linear and Logistic Regression

It is use to reduce the multicolinearity of the independent variable, because overfitting

Here we reduce Training Accuracy and Increase Testing Accuracy by lowering the m value tha
Here we just add this term, lambda*(slope^2) in linear and logistic regression cost functi
we use lambda*|slope| instead of square. For more features we prefer L1 i.e Lasso.

'''
```