Interview questions on LR

Q1. What is the loss function used in linear regression?
Ans→ The loss function used in linear regression is OLS.

Q2. What are the basic assumptions of LR?
Ans→ There are five basic assumptions of linear regressions—
(a) There should be linear relationship between dependent and independent variables.
(b) Little or no multicollinearity.
(c) The residuals should be homoscedastic (should have the same variance) and should be randomly distributed along the best fit line/hyperplane.
(d) The residuals should be normally distributed.
(e) There should be very little or no autocorrelation.

Q3. Explain the gradient descent algorithm with respect to linear regression?
Ans→ Gradient descent is a first-order optimization algorithm. In LR, this algorithm is used to find the values of the θ_i (estimators/coefficients) corresponding to the optimized value of the cost function.

Q4. List down the metrics that could be used to evaluate a LR model?
Ans→ Various evaluation metrics that can be used for a LR model are — (i) MSE (ii) RMSE (iii) MAE (iv) MAPE (v) R^2 (vi) adj R^2

Q5. For a linear regression model, how do we interpret a Q-Q plot?
Ans→ The Q-Q plot represents a graphical plotting of the quantiles of two distribution with respect to each other. In simple words, we plot quantiles in the Q-Q plot which is used to check normality of errors.
Whenever we interpret a Q-Q plot, we should concentrate on the '$y=x$' line which corresponds to a normal distribution. It implies that each of the distributions has the same quantiles. In case we see a deviation from this line, one of the distributions can be skewed when compared to other i.e., normal distribution.

Q6. In linear regression, what is the value of the sum of residuals for a given dataset?

Ans→ The sum of residuals in a linear regression model is 0 since it assumes that the errors (residuals) are normally distributed with an expected value or mean equal to 0.

Q7. Which evaluation metric should we prefer to use for a dataset having a lot of outliers in it?

Ans→ Mean Absolute Error (MAE) is preferred when we have too many outliers present in the dataset because MAE is robust to outliers whereas MSE and RMSE are very susceptible to outliers and these start penalizing the outliers by squarring the error terms.

Q.8 What is R-squared and adj R-squared?

Ans→ Refer page no. 8-9.

Q.9 What is multicollinearity? How to detect it?

Ans→ Refer page 3.

Q.10→ What is heteroscedasticity? How to detect it?

Ans→ Refer page 4. It can be detected using graphs or some statistical tests such as Breush-Pagan test and NCV test.

Q.11 What is VIF and tolerance?

Ans→ Refer page 3.

Q.12 How is hypothesis testing used in linear regression algorithm?

Ans→ We can carry out hypothesis testing in linear regression for the following purposes—

(a) To check whether an independent variable (predictor) is significant or not for the prediction of target variable. Two common methods for this are—

(i) by using p-values → If the p-value of a particular independent variable is greater than a certain threshold (usually 0.05), then that independent variable is insignificant for the prediction of the ~~the~~ target variable

(ii) by checking the values of the regression coefficient →
If the value of the regression coefficient corresponding to a particular independent variable is 0, then that variable is insignificant for the prediction of the dependent variable and has no linear relationship with it.

(b) To verify whether calculated ~~calculated~~ regression coefficients are good estimators or not of the actual coefficients.

Q.13 Is it possible to apply linear regression for time-series analysis?

Ans→ Yes, we can use linear regression to do a time-series analysis, but the results could be not promising, hence it is not advisable to do so.

The reasons behind not advising LR to do time series analysis are —

* Time series data is mostly used for the prediction of the future, but in contrast linear regression generally seldom gives good results for future predictions as it is basically not meant for extrapolation.

** Moreover, time-series data have a pattern such as peak hours, festive seasons, etc. which could most likely be treated as outliers in the linear regression analysis.

Q.14 Why do we sum of squarred residuals (SSR) instead of the sum of additional absolute errors (SAE) in linear regression. Explain it with examples.

Ans→ The major reason of choosing (SSR) over (SAE) comes down to mathematical properties and underlying assumptions about the data. Some of the common reasons are —

(a) Differentiability: SSR leads to a diffrentiable loss function, while SAE does not. This makes optimization easier using techniques like gradient descent.

(b) Robustness: While SAE is less sensitive to outliers compared to RSS, it gives equal weight to all errors regardless of their magnitude. In some cases, this might not be desirable, especially when we want that larger errors should be penalized more heavily, RSS/SSR by squarring the errors penalizes larger errors more than smaller errors which is more beneficial in many cases.

(c) Computational efficiency: SAE can be computationally more expensive for optimizating purpose whereas RSS has less expensive closed form solutions available.

Q.15 What can we do if we overfit a data using linear regression?

Ans→ Firstly, overfitting means that our model has learned the training data very well that the training error is very less but when an unseen data point comes the testing error is high conversly variance is high.
To solve this problem we can use regularization techniques like L1 and L2 regularization which adds a penalty term to the normal loss function of LR and ~~penalizes~~ shrinks the coefficients.

Q.16 What is the difference between L1 and L2 regularization?

Ans→ Refer page 10 & 11. L2 regularization shrinks the coefficients so that it makes coefficient of all the parameters/features more uniform and thus one parameter doesn't dominate others whereas L1 regularization promotes sparsity by eleminating/shrinking the coefficient of irrelevant features to 0 thus those features will be dropped and it can be used for feature selection as well.

Q.17 Explain the analytical/normal/closed form solution in linear regression in detail and where it can be used?

Ans→ In contrast to gradient descent, we can use closed form solution as well to find the coefficients which minimizes the cost function.
The closed form solution is computed in one-go and calculates the exact coefficients.

The closed form solution has a mathematical expression given as —

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

This method of calculating coefficients should be preferred for smaller datasets because finding the inverse of a matrix can be more computationally expensive also sometime inverse of a matrix does not exist when the determinant of the matrix is 0.

Q.18 When should it be preferred to use the gradient descent method instead of normal/closed form method to evaluate the best coefficients for linear regression?

Ans→

Gradient Descent	Closed form
* It needs hyperparameter tuning to find the best value of α (learning rate).	* No need of hyperparameter tuning.
** It is an iterative process.	** Solution is computed in one go
*** Time complexity → $O(\underbrace{k}_{\text{no. of iterations}} n^2)$	*** Time complexity → $O(n^3)$ due to evaluation of $X^T X$.
**** It is preferred when data points are large.	**** It is preferable in case of small dataset.