We at The Data Monk hold the vision to make sure everyone in the IT industry has an equal stand to work in an open domain such as analytics. Analytics is one domain where there is no formal under-graduation degree and which is achievable to anyone and everyone in the World.

We are a team of 30+ mentors who have worked in various product-based companies in India and abroad, and we have come up with this idea to provide study materials directed to help you crack any analytics interview.

Every one of us has been interviewing for at least the last 6 to 8 years for different positions like Data Scientist, Data Analysts, Business Analysts, Product Analysts, Data Engineers, and other senior roles. We understand the gap between having good knowledge and converting an interview to a top product-based company.

Rest assured that if you follow our different mediums like our blog cum questions-answer portal www.TheDataMonk.com , our youtube channel - The Data Monk, and our e-books, then you will have a very strong candidature in whichever interview you participate in.

There are many blogs that provide free study materials or questions on different analytical tools and technologies, but we concentrate mostly on the questions which are asked in an interview. We have a set of 100+ books which are available both on Amazon and on The Data Monk e-shop page

We would recommend you to explore our website, youtube channel, and e-books to understand the type of questions covered in our articles. We went for the question-answer approach both on our website as well as our e-books just because we feel that the best way to go from beginner to advance level is by practicing a lot of questions on the topic.

We have launched a series of 50 e-books on our website on all the popular as well as niche topics. Our range of material ranges from SQL, Python, and Machine Learning algorithms to ANN, CNN, PCA, etc.

We are constantly working on our product and will keep on updating it. It is very necessary to go through all the questions present in this book.

Give a rating to the book on Amazon, do provide your feedback and if you want to help us grow then please subscribe to our Youtube channel.

SVM Interview Questions

Q1. What is the full form of SVM?

a. Support vending machinery
b. Support vector machine
c. Sequence vector machine
d. Sequence vector mode

Answer: (b) Support vector machine

Q2. The effectiveness of an SVM depends upon.

a. Kernel Parameters
b. Selection of Kernel
c. Soft Margin Parameter C
d. All of the mentioned

Answer: (d) All of the mentioned

Q3. Which statement are true about SVM?

Statement 1: Kernel function map low dimensional data to high dimensional space

Statement 2: It's a similarity function

a. Both the options are right
b. Statement 1 is true
c. Statement 2 is true
d. Both are false

Answer: (a) Both the options are right

Q4. Support vectors are the data points that lie closest to the decision surface.

a. Partially true
b. True
c. False
d. Invalid statement

Answer: (b) True

Q5. What do you mean by generalization error in terms of SVM?

a. How far the hyperplane is from support vectors.
b. How accurately the SVM can predict the outcome for unseen data.

c. The threshold amount of error in an SVM.
d. None of the above

Answer: (b) How accurately the SVM can predict the outcome for unseen data.

Q6. What would happen if you use very small C(C~0)?

a. Data will be correctly classified
b. Can't say
c. Misclassification would happen
d. None of these

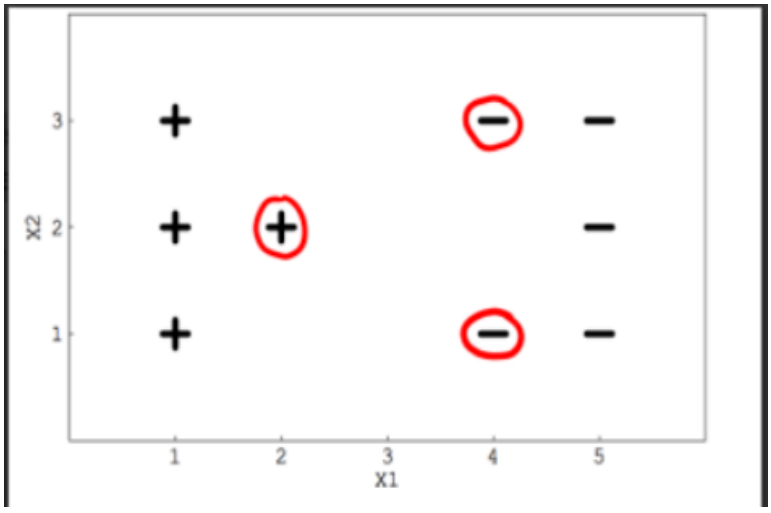
Answer: (c) Misclassification would happen

Q7. The SVM's are less effective when:

a. The data is linearly separable.
b. The data is clean and ready to use.
c. The data is noisy and contain overlapping points.
d. All of the above

Answer: (c) The data is noisy and contain overlapping points.

Q8. Suppose you are using a Linear SVM classifier with 2 class classification problem. Now you have been given the following data in which some points are circled red that are representing support vectors.



If you remove one of the red points from the data (It can be any of three). Does the decision boundary will change?

a. No
b. Yes

c. Partially YEs
d. Invalid

Answer: (b) Yes

Q9. Suppose you are dealing with 4 class classification problem and you want to train a SVM model on the data for that you are using One-vs-all method. How many times we need to train our SVM model in such cases?

a. 5
b. 7
c. 4
d. 3

Answer: (c) 4

Q10. What do you mean by hard margin?

a. SVM allows very low error in classification
b. SVM allows high amount of error in classification.
c. None of the above
d. Both a and b

Answer: (a) SVM allows very low error in classification

Q11. When the C parameter is set to infinite, which of the following holds true?

a. The soft-margin classifier will separate the data
b. The optimal hyperplane if exists, will be the one that completely separates the data
c. Both a and c
d. None of the above

Answer: (b) The optimal hyperplane if exists, will be the one that completely separates the data

Q12. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify?

a. The model would not be affected by distance of points from hyperplane for modeling
b. The model would consider even far away points from hyperplane for modeling
c. The model would consider only the points close to the hyperplane for modeling
d. All the above options are correct

Answer: (c) The model would consider only the points close to the hyperplane for modeling.

Q13. What do you mean by cost parameter in SVM?

a. The tradeoff between misclassification and simplicity of the model
b. The kernel to be used
c. The number of cross-validations to be made
d. None of the above

Answer: (a) The tradeoff between misclassification and simplicity of the model

Q14. We usually use feature normalization before using the Gaussian kernel in SVM. What is true about feature normalization?

a. Sometimes, feature normalization is not feasible in case of categorical variables.
b. Feature normalization always helps when we use Gaussian kernel in SVM.
c. We do feature normalization so that new feature will dominate other.
d. All the above
e. Option a and c

Answer: (e) Option a and c that is Sometimes, feature normalization is not feasible in case of categorical variables, we do feature normalization so that new feature will dominate other.

Q15. What are real world examples of SVM?

a. Speech Recognition
b. Text Classification
c. Facial Expression Classification
d. Inverse Geosounding Problem
e. All the above

Answer: (e) All the above

Q16. Explain SVM in simple words.

In simple words SVM is an algorithm creates a line or a hyperplane which separates the data into classes. Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems.

Q17. Explain different types of SVM.

SVM are of two types:

1. Linear SVM: If a dataset can be classified into two classes by using a one straight line, then such data is called as linearly distinguishable data, and classifier is used called as Linear SVM classifier.
2. Non-linear SVM: If a dataset cannot be classified by using one straight line, then such data is known as non-linear data and classifier used is called as Non-linear SVM classifier.

Q18. Explain hyperplane of SVM in simple words.

A hyperplane in an n-dimensional Euclidean space is a space that divides the space into two disconnected parts or divides a plane into two parts which may or may not be look same.

Q19. What is margin and optimal hyperplane in SVM?

The distance between the vectors and the hyperplane is known as margin and the hyperplane with maximum margin is called the optimal hyperplane.

Q20. Does SVM give any probabilistic output?

SVMs do not directly provide probability estimates, it is calculated using an expensive five-fold cross-validation.

Q21. What is the geometric intuition behind SVM?

The basic aim behind support vector machine is to make a hyperplane which can divide a plan into two classes or two parts distinguished by a line. SVM draws a hyperplane parallel to the actual hyperplane intersecting with the first point of class A (also known as Support Vectors) and another hyperplane parallel to the actual hyperplane intersecting with the first point of class B.

Q22. How would explain Convex Hull in light of SVMs?

We simply build a convex hull for class A and class B and draw a perpendicular on the shortest distance between the closest points of both these hulls.

Convex Hull in SVM is used to efficiently remove insignificant data and thereby reduce the training time of SVM.

Q23. What is a Kernal in SVM?

Kernels are simply adds more features to the data for making it linearly separable. Two popular kernels are the polynomial kernel and the Gaussian Radial Basis Function, or RBF, kernel.

Q24. What are different types of Kernal in SVM?

1. Gaussian kernel
2. Polynomial kernel
3. Laplace RBF kernel.
4. Gaussian radial basis function (RBF)
5. Sigmoid kernel
6. Hyperbolic tangent kernel.
7. Bessel function of the first kind Kernel

Q25. What is the kernel trick polynomial?

The kernel trick allows you to save time/space and compute dot products in an n dimensional space. You want to work with degree 2 polynomial features, ø(x). Then, your dot product will be operate using vectors in a space of dimensionality n(n+1)/2.

Q26. What is Gaussian kernel?

One of the most popular kernel in SVM is the Gaussian kernel, Gaussian kernel is represented as ||X1 — X2 || = Euclidean distance between X1 and X2. It is more often used than polynomial kernels when learning from nonlinear datasets.

Q27. Is SVM supervised or unsupervised machine learning algorithm?

SVM that is Support vector machines is a supervised machine learning algorithm which works both on regression problems and classification problems.

Q28. Explain about SVM Regression?

The SVR that is Support vector regression has only few differences as compared to Support vector machine but it mostly uses same principles. We can use SVR for working with continuous Values instead of Classification which is SVM.

Q29. SVM being a large margin classifier, is it influenced by outliers?

Yes, if C is large SVM is it influenced by outliers.

Q30. What is the difference between logistic regression and SVM without a kernel?

Logistics regression and SVM are mostly same but the major difference between them is that SVM is much more efficient and has good optimization packages.

Logistic Regression algorithm is used for solving classification problems.

Support vector machine is a model used for both classification and regression.

Logistic regression is based on statistical approach.

Support vector machine is based on geometrical properties of data.

The example of logistic regression is Cancer Detection problem can be solved by logistic regression.

The example of support vector machine is Image classification and HAndwritting recognition.

Q31. What does RBF kernel do?

Radial Basis Kernel is a kernel function that is used in machine learning to find a non-linear classifier or regression line.

Q32. What are Support Vectors in SVMs?

Support vectors are those instances that are located on the margin itself. For SVMS, the decision boundary is entirely determined by using only the support vectors.

Q33. What are soft margins and hard margins in SVM?

A hard margin means that an SVM is very rigid in classification and tries to work extremely well in the training set, causing overfitting.

Support Vector Machine is trying to find the line that maximizes the margin (think of a street), which is the distance between those closest dots to the line. This is called the Soft Margin.

Q34. State the properties of Hinge loss function?

The hinge loss function is defined as max(0,1-t)

This function is not differentiable at t=1

The functions value is zero when the value of t is greater than or equal to 1 that is t>=1

The functions derivative is equal to -1 if t<1 and 0 if t>1.

Q35. Is SVM sensitive to the Feature Scaling?

Yes, SVMs are sensitive to feature scaling as it takes input data to find the margins around hyperplanes and gets biased for the variance in high values.

Q36. What are the real world applications of SVM?

Facial Expression Classification.

Speech Recognition

Text Classification.

Data Classification using SSVM

Stenography Detection in Digital Images

Cancer Diagnosis and Prognosis

Q36. Which is better SVM or KNN?

KNN is better than SVM, SVM take cares of outliers better than KNN. If training data is much larger than no. of features(m>>n), then KNN is better than SVM.

Q37. Why Gaussian kernel is used in SVM?

In SVM Gaussian kernel is used for solving nonlinear problems such as X-OR in higher dimensional where linear separation is not possible. Gaussian kernel also gives good linear separation in higher dimension for many nonlinear problems.

Q38. Give some situations where you will use an SVM over a Random Forest Machine Learning algorithm and vice-versa?

These are the situations where we can use an SVM over a random forest machine learning algorithm, the factors are:

The hyperparameters which you choose.

The factors on how you evaluate your model overall.

Also it can be checked by the distribution of the data

Also by the input scale of feature.

Q39. What is generalization error in terms of the SVM?

Generalization error in terms of SVM is basically the out-of-sample error which is the proportion of how exactly a model can predict values for past unseen data.

Q40. What is the role of C in SVM?

The C parameter tells the SVM optimization that how much you want to avoid misclassifying each training example.

Q41. How is SVM different from other classification method such as KNN and logistic regression?

SVM is based on geometrical properties of the data while logistic regression is based on statistical approaches. Logistic regression works with already identified independent variables

whereas SVM works well with unstructured and semi-structured data like text and images.

Q42. What are the basic steps to build a project based on SVM?

1. Firstly you need to import all the libraries which you need.
2. After importing libraries, you must load your dataset which you will use for the project.
3. Split the data in training and testing subsets.
4. Here is the main step where we apply the Support vector machine algorithm and build the model.
5. In the last step you can check the accuracy on test data and see the result.

Q43. Explain linear kernel in SVM.

A linear kernel can be used as a normal dot product between any two given observations. The product between the two vectors is the sum of the multiplication of each pair of input values.

Q44. List out the advantages of using SVM.

Uses a subset of training points in the decision function that makes it memory well oraganized.

SVM's are very good when we have no idea on the data.

Different kernel functions can be described for the decision function that also makes it multiskilled.

Works well with even unstructured and semi structured data like text, Images and trees.

SVM is always compared with ANN. When compared to ANN models, SVMs give better results.

Effective in high dimensional spaces.

Q45. List the disadvantage of using SVM.

Everytime choosing a best kernel is not at all easy.

The SVM hyper parameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact.

It takes long training time for large datasets.

SVMs do not directly provide probability estimates, these are calculated using five-fold cross-validation.

Questions on Linear Regression

Q46. Linear Regression is a supervised machine learning algorithm.

a. Partially True
b. True
c. False
d. Statement is invalid

Answer: (b) True

Q47. Linear Regression is mainly used for Regression.

a. True
b. False

Answer: True

Q48. Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable?

a. Mean-Squared-Error
b. Logloss
c. AUC-ROC
d. Accuracy

Answer: (a) Mean-Squared-Error

Q49. If Linear regression model perfectly first i.e., train error is zero, then _____ complete the following sentence.

a. Couldn't comment on Test error
b. Test error is also always zero
c. Test error is equal to Train error

d. Test error is non zero

Answer: (a) Couldn't comment on Test error

Q50. Which of the following metrics can be used for evaluating regression models?

i) R Squared

ii) Adjusted R Squared

iii) F Statistics

iv) RMSE / MSE / MAE

a. i and ii
b. ii, iii and iv
c. i, ii, iii and iv
d. ii and iv

Answer: (c) i, ii, iii and iv

Q51. Function used for linear regression in R is _____

a. regression.linear(formula, data)
b. mlr(formula, data)
c. lrs(formula, data)
d. lm(formula, data)

Answer: (d) lm(formula, data)

Q52. In practice, Line of best fit or regression line is found when _____

a. Mean of the square of residuals ($\sum$ (Y-h(X))2) is minimum
b. Sum of the square of residuals ($\sum$ (Y-h(X))2) is minimum
c. Sum of residuals ($\sum$(Y – h(X))) is minimum
d. Mean of the absolute value of residuals ($\sum$|Y-h(X)|) is minimum

Answer: (b) Sum of the square of residuals (∑ (Y-h(X))2) is minimum

Q53. Which of the following is true about Residuals?

a. Higher is better
b. Lower is better
c. A or B depend on the situation
d. None of these

Answer: (b) Lower is better

Q54. The vertical spread of the data points about the regression line is measured by what?

a. t-ratio
b. standard error of estimate
c. regression coefficient
d. y-intercept
e. homoscedasticity coefficient

Answer: (d) y-intercept

Q55. Fill in the blanks with appropriate answer.
The coefficient of determination can take on values between _____ and _____, inclusive.
Answer: 0 and 1

Q56. What does linear regression exactly mean?
Linear regression is a machine learning algorithm which finds the best linear-fit relationship on any given data, between independent and dependent variables.

Q57. What are outliers? How do you detect and list the disadvantage of it?
Outlier is the data which is totally different from the other data or in other words we can say that it is a value that escapes normality and can cause error in the result.

We can detect outliers by the use of box plot. Box plots use the median and the lower and upper quartiles.

If we don't want outliers in our data we can use. Winsorization method, in this method we can trim the data set, but replace outliers with the nearest "good" data,

as opposed to truncating them completely or else we can replace outliers with the mean or median.

The disadvantage would be outliers can have a disproportionate effect on statistical results, such as the mean, which can result in misleading interpretations.

Q58. What is difference between simple linear and multiple linear regressions?

- Linear Regression: It establishes the relationship between two variables using a straight line. Linear regression attempts to draw a line that comes closest to the data by finding the slope and intercept that define the line and minimize regression errors.
- Multiple regressions: Multiple regressions are based on the assumption that there is a linear relationship between both the dependent and independent variables. It also assumes no major correlation between the independent variables.
- Linear Regression: Simple linear regression has only one x and one y variable.
- Multiple Regression: Multiple linear regression has one y and two or more x variables.
- Linear Regression: In simple linear regression we predict rent based on square feet alone.
- Multiple Regression: In multiple linear regression we predict rent based on square feet and age of the building.

Q59. In regression analysis, which of the statements is true?

Statement 1: The Mean of residuals is less than Zero at all times.
Statement 2: The Mean of residuals is more than Zero at all times.
Statement 3: The mean of residuals is always equal to Zero.
Statement 4: You do not have any such rule for residuals.
Answer: Among all the statements, Statement 3: The mean of residuals is always equal to Zero is correct.

Q60. What is the importance of the F-test in a linear model?

The F-test is an important one in the sense that it tests the principle of the model. When you repeat the model to upgrade the exactness with the changes, the F-test proves its usefulness in understanding the effect of the all over regression.

Q61. What is a residual? How is it determined?

The other term of residual is error, residual is the difference between the predicted y value and the actual y value. Residual can be positive or negative. If residuals are always 0, then your model has a Perfect R square i.e. 1.
Residual = Actual y value – Predicted y value.

Q62. What is P-Value and how do you find P value in linear regression?

Value is a probability score that is used in statistical tests to establish the statistical significance of an observed effect. Or simple regression, the p-value is determined using a t distribution with n − 2 degrees of freedom (df), which is written as t n − 2 and is calculated as 2 × area past |t| under a t n − 2 curve.

Q63. How do you know that linear regression is suitable for any given data?

To see if linear regression is suitable for any given data, a scatter plot can be used. If the relationship looks linear, we can go for a linear model. Plotting the scatter plots is easy in case of simple or univariate linear regression. But in case of multivariate linear regression, two-dimensional pairwise scatter plots, rotating plots, and dynamic graphs can be plotted.

Q64. What is robust regression and when should I use robust regression?

Robust regression is a form of regression analysis designed to overcome some limitations of traditional parametric and non-parametric methods.

Robust regression is an iterative procedure that seeks to identify outliers and minimize their impact on the coefficient estimates.

Q65. What are the Four Assumptions of Linear Regression?

1. Linear relationship: There exists a linear relationship between the independent variable, x, and the dependent variable, y.

2. Independence: The residuals are independent. In particular, there is no correlation between consecutive residuals in time series data.

3. Homoscedasticity: The residuals have constant variance at every level of x.

4. Normality: The residuals of the model are normally distributed.

Q66. How is Total Degrees of Freedom for Linear Regression is calculated?

The total degrees of freedom for the linear regression model is taken as the sum of the model degrees of freedom plus the model error degrees of freedom. Generally, the degrees of freedom is equal to the number of rows of training data used to fit the model.

Q67. How is linear regression related to correlation?
Correlation quantifies the direction and strength of the relationship between two numeric variables, X and Y, and always lies between -1.0 and 1.0. Simple linear regression relates X to Y through an equation of the form Y = a + bX.

Q68. Can we use linear regression for time series analysis?
One can use linear regression for time series analysis, but the results are not sure or promising.

Q69. What happens when p value for f test is lower than alpha?
When p value for f test is lower than alpha (which is usually .05 if nothing else is specified), then we reject H0.

Q70. What do you conclude from above question?
We conclude that: We have the evidence that at least one of the [x variables] has a significant relationship with the [y variable].

Q71. In regression forecasting, what do we mean when we say that there is linearity in a set of data?
If there is linearity, then it helps in predicting the future values of a variable with the help of the past values. It indicates that the mean values of a dependent variable for every one unit variation in its independent variables tend to be in a straight line, and in other words, it shows that it creates a straight-line relationship.

Q72. What is Q-Q plot in a linear regression model?
The Q-Q plot is a graphical plotting of the quantiles of two distributions with respect to each other. In other words, you plot quantiles against quantiles.

Q73. List the advantages of linear regression.

1. Simple Implementation: Linear Regression is a very simple algorithm that can be implemented very easily to give satisfactory results.
2. Performance on linearly separable datasets: Linear regression fits linearly separable datasets almost perfectly and is often used to find the nature of the relationship between variables.
3. Overfitting can be reduced by regularization: Overfitting is a situation that arises when a machine learning model fits a dataset very closely and hence captures the noisy data as well. Regularization is a technique that can be

easily implemented and is capable of effectively reducing the complexity of a function so as to reduce the risk of overfitting.

Q74. List the disadvantages of Linear Regression.

1. Linear Regression only looks at the mean of the dependent variable.
2. Linear Regression is limited to linear relationships.
3. Linear Regression is sensitive to outliers.
4. Data must be independent.

Q75. Who uses linear regression?

Linear regression is commonly used for predictive analysis and modeling.

Q76. What is one real life example of when regression analysis is used?

A simple linear regression real life example could mean you finding a relationship between the revenue and temperature, with a sample size for revenue as the dependent variable. In case of multiple variable regression, you can find the relationship between temperature, pricing and number of workers to the revenue.

Q77. If you have only one independent variable, how many coefficients will you require to estimate in a simple linear regression model?
Two coefficients will you require to estimate in a simple linear regression model.

Q78. Which plot is best suited to test the linear relationship of independent and dependent continuous variables?
Scatter Plot is best suited to test the linear relationship of independent and dependent continuous variables.

Q79. What are the columns in the coefficient table?

The coefficient table contains the variable name, coefficient, standard error and p-value.

Q80. What is OLS?

OLS stands for ordinary least square, it measures the error between the actual Y and predicted Y, lower the error, better is the model.

OLS estimators minimize the sum of the squared errors (a difference between observed values and predicted values).

Q81. What is the relationship between R-Squared and Adjusted R-Squared?

Adj R-Sq is always lower than the R-Sq, the relationship between them is mentioned below:

Adj R-Sq = 1 – ((1-RSq) * (n-1) / (n-p-1))

Where n is the number of observations

p is the number of variables

Q82. What happens if the linear regression violates any of its assumptions?

If the regression assumptions are violated, then performing regression analysis will give incorrect results.

Q83. How to identify if the model is overfitting or underfitting?

Overfit model performs good (high accuracy) on training and bad (low accuracy) on test.

Underfit model performs bad (low accuracy) on training and bad (low accuracy) on test.

A good model performs good (high accuracy) on training and good (high accuracy) on test.

Q83. What is adjusted R-Squared?

Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. As we add more variables the descriptive power of the model may increase. Adjusted R-Squared punish the model for the number of variables that are used in the model.

Q84. How do you explain Multicollinearity and when does this happen?

Multicollinearity is the circumstance of high intercorrelations of two or more independent variables in a multiple regression model. Multicollinearity occurs when independent variables in the regression model are highly correlated to each other. It makes it worst to interpret of model and also creates an overfitting problem.

Q85. List some other regression methods.

- Gaussian regression
- Kernel regression
- Bayesian regression

- Generalized least squares

Q86. What is the relationship between R-Squared and Adjusted R-Squared?

Adj R-Sq = 1 – ((1-RSq) * (n-1) / (n-p-1))

R-Sq is always more than Adj R-Sq

Q87. What does no endogeneity mean?

No endogeneity means no relationship between x and ε, it occurs because we have omitted an important predictor from the model.

Q88. What is the main purpose of Linear Regression and Logistic Regression?

The purpose of Linear Regression is to find the best-fitted line while Logistic regression is one step ahead and fitting the line values to the sigmoid curve.

Q89. How do you interpret linear regression in Excel?

1. On the Data tab, in the Analysis group, click the Data Analysis button.
2. Select Regression and click OK.
3. In the Regression dialog box, configure the following settings: Select the Input Y range, which is your dependent variable.
4. Click OK and observe the regression analysis output created by Excel.

Q90. Which is better logistic regression or linear regression?

Logistic regression is better because it has Good accuracy for many simple data sets and it performs well when the dataset is linearly separable, it requires average or no multicollinearity between independent variables.

Q91. What is the common problem with linear regression?

Linear regression assumes that the data are independent. That means that the scores of one subject (such as a person) have nothing to do with those of another.

Q92. What will happen when you fit degree 4 polynomial in linear regression?

Degree 4 will be more complex(overfit the data) than the degree 3 model so it will again perfectly fit the data. In such case training error will be zero but test error may not be zero.

Q93. How is RMSE is calculated without formula?

- Squaring the residuals.
- Finding the average of the residuals.
- Taking the square root of the result.

Q94. Write the formula for RMSE.

$$RMSE = \sqrt{\overline{(f-o)^2}}$$

Where:

- f = forecasts (expected values or unknown results),
- o = observed values (known results).

Q95. Is a higher or lower RMSE better?

Lower values of RMSE indicate better fit. RMSE is a good measure of how accurately the model predicts the response, and it is the most important criterion for fit if the main purpose of the model is prediction.

Q96. How do you calculate MSE?

1. Find the regression line.
2. Insert your X values into the linear regression equation to find the new Y values (Y').
3. Subtract the new Y value from the original to get the error.
4. Square the errors.
5. Add up the errors
6. Find the mean.