1. Data Science is a field that involves the extraction of insights and knowledge from data. It combines several disciplines such as statistics, computer science, and domain knowledge. Supervised learning and unsupervised learning are two categories of machine learning algorithms. Supervised learning algorithms use labeled data to learn patterns and make predictions, while unsupervised learning algorithms work with unlabeled data and aim to discover patterns and structures in the data.
2. Logistic regression is a statistical method used for binary classification problems. It models the relationship between a dependent binary variable and one or more independent variables.
3. Multiclass classification problems involve predicting the outcome of a categorical variable with more than two classes. Logistic regression can be extended to handle multiclass problems by using techniques such as one-vs-rest or softmax regression.
4. Linear regression is a statistical method used to model the relationship between a continuous dependent variable and one or more independent variables, while logistic regression is used to model the relationship between a binary dependent variable and one or more independent variables.
5. Logistic regression is popular because it is simple, interpretable, and can be used for a variety of classification problems. It also has a probabilistic interpretation, which makes it useful for assessing the uncertainty of predictions.
6. The formula for the logistic regression function is:

P(Y=1|X) = 1 / (1 + e^-(b0 + b1X1 + ... + bnXn))

where P(Y=1|X) is the probability of the dependent variable Y being equal to 1 given the independent variables X1, X2, ..., Xn, and b0, b1, ..., bn are the coefficients of the model.

1. The assumptions made in logistic regression include linearity, independence of errors, homoscedasticity, and absence of multicollinearity.
2. Logistic regression is called regression because it estimates the relationship between the dependent variable and the independent variables. The output of the logistic regression model is a continuous probability score, which can be used to classify observations into different categories.
3. The general intuition behind logistic regression is to find a mathematical function that can map the values of the independent variables to the probability of the dependent variable taking on a particular value.
4. The sigmoid function is a mathematical function that is used in logistic regression to map the output of the linear regression function to a probability score. It is an S-shaped curve that ranges from 0 to 1 and has the property of being monotonically increasing.
5. Gradient descent is an optimization algorithm used to minimize the cost function in logistic regression. It works by iteratively adjusting the model parameters in the direction of steepest descent of the cost function.
6. Mean Square Error (MSE) cannot be used as a cost function for logistic regression because the output of the logistic regression model is a probability score, which is not continuous and does not have a normal distribution.
7. The Confusion Matrix is a table that is used to evaluate the performance of a classification model. It shows the number of true positives, true negatives, false positives, and false negatives.
8. False positives are cases where the model predicts a positive outcome when the actual outcome is negative, while false negatives are cases where the model predicts a negative outcome when the actual outcome is positive.
9. The true positive rate (TPR) is the proportion of positive cases that are correctly identified by the model, while the true negative rate (TNR) is the proportion of negative cases that are correctly identified by the model.
10. The false-positive rate (FPR) is the proportion of negative instances that are incorrectly predicted as positive by a classification model. The false-negative rate (FNR) is the proportion of positive instances that are incorrectly predicted as negative by a classification model.
11. Precision and recall are evaluation metrics used in binary classification problems. Precision is the proportion of true positive predictions among all positive predictions made by the model. Recall is the proportion of true positive predictions among all actual positive instances in the dataset. These metrics are important in model evaluation because they provide insight into the trade-off between making correct predictions and avoiding incorrect predictions.
12. The precision-recall curve is a graphical representation of the trade-off between precision and recall for different classification thresholds. This curve is useful for comparing and selecting different models based on their precision-recall performance.
13. The F1-score is a single evaluation metric that combines precision and recall into a single value. It is the harmonic mean of precision and recall and provides a balance between the two metrics. The F1-score is important because it provides a single value that can be used to compare and select different models.
14. Classification accuracy alone is not considered a good measure in logistic regression because it may be biased towards the majority class in imbalanced datasets. Additionally, it does not provide insight into the trade-off between making correct predictions and avoiding incorrect predictions.
15. Accuracy can be calculated using a confusion matrix by summing the diagonal elements (true positives and true negatives) and dividing by the total number of instances.
16. Sensitivity is the proportion of true positive predictions among all actual positive instances in the dataset.
17. Specificity is the proportion of true negative predictions among all actual negative instances in the dataset.
18. The ROC curve is a graphical representation of the trade-off between sensitivity and specificity for different classification thresholds. The area under the ROC curve (AUC) is a single evaluation metric that provides a measure of overall performance across all classification thresholds.
19. The bias-variance tradeoff is a fundamental concept in machine learning that refers to the trade-off between model complexity and generalization performance. Bias refers to the error that results from a model's inability to capture the true relationship between the input features and the target variable. Variance refers to the error that results from a model's sensitivity to noise in the training data.
20. Overfitting occurs when a model is overly complex and fits the noise in the training data, resulting in poor generalization performance on new data.
21. Underfitting occurs when a model is too simple and is not able to capture the true relationship between the input features and the target variable, resulting in poor performance on both the training and test data.
22. To deal with overfitting in machine learning, we can use techniques such as regularization, early stopping, or using a simpler model with fewer features. Underfitting, on the other hand, can be addressed by increasing the model complexity, adding more features, or using a more powerful algorithm.
23. In the bulls-eye diagram, the center represents the true function that we want to approximate, while the outer rings represent models of increasing complexity. Bias refers to the error caused by the model's inability to capture the true function, while variance refers to the error caused by the model's sensitivity to noise in the training data. A good model should have low bias and low variance, which is represented by the sweet spot in the middle of the diagram.
24. Advantages of logistic regression include its simplicity, interpretability, and ability to handle nonlinear relationships between predictors and outcomes.

Disadvantages include its assumption of linearity between predictors and the log odds of the outcome, its inability to handle complex interactions between predictors, and its susceptibility to overfitting when the number of predictors is large compared to the number of observations.

**Logistic Regression:**

Logistic regression is a classification algorithm used to find the probability of event success and event failure.

Logistic regression is a generalized linear model and it uses the same basic formula of linear regression.

Y = sigmoid(b0 + b1X1 + b2X2 + ... + bnXn)

| **Advantages** | **Disadvantages** |
| --- | --- |
| Easy to implement and interpret yet efficient in training | Overfits on high dimensional data |
| The predicted parameters give inference about the importance of each feature | Non linear problems can't be solved with logistic regression since it has a linear decision surface |
| Performs well on low-dimensional data. | Assumes linearity between dependent and independent variables. |
| Very efficient when the dataset has features that are linearly separable. | Fails to capture complex relationships. |
| Outputs well-calibrated probabilities along with classification results. | Only important and relevant features should be used otherwise model's predictive value will degrade. |

* Logistic Regression algorithm can easily be extended to multi-class classification using a Softmax Classifier, this is known as Multinomial Logistic Regression.