Questions

IMPORTANT QUESTIONS:

1. What is regression?

Ans: Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable.

2. Which method is used to describe the best fit line in regression?

Ans: Least-square method

Q.3- What type of data is used in the Regression output?

Ans: Continuous data. Predicts a continuous dependent variable based on one or more independent variables.

Q.4. What are the types of Regression?

Ans: Linear Regression: Linear regression model is used to test the relationship between two variables in the form of an equation.

Logistic Regression:

Q.5. Write the model equation of Regression having multiple features.

Any general **linear equation** with **multiple features** can be written as :

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \ldots\ldots\ldots + a_nx_n$$

$$x_i \text{ are features}$$
$$a_i \text{ are model parameters or coefficients}$$
$$y \text{ is target variable}$$

Q.6 What is Global Minimum ?

Ans: The optimal point at which the cost function is minimum in a machine learning model.

Q.7. Gradient Descent is an Optimization algorithm/ Regression algorithm.

Ans: Optimisation algorithm

Q.8 How Gradient Descent is used in Neural Network?

Ans: Gradient Descent is an optimisation algorithm. Gradient descent is used to find the optimal value of parameters by adjusting the parameters such as weights and biases of the neural network to minimise the loss function.

Q.9 Gradient is also known as the _____ of the tangent.

Ans: Slope

Q.10 How to calculate Gradient of an equation?

Ans: Partial derivative of both variables x and y

Q.11 What is Gradient vector?

Ans: In a higher dimensional function, the gradient is a vector made up of all the partial derivatives of the function.

Q.12 Name the method that updates the parameters/ variables of the cost function in the direction of negative gradient in an iterative mannet to reach the minimum of the cost function.

Ans: Gradient Descent

Q.13 Cost function is also known as _____ function.

Ans: Loss function or objective function

Q.14 What is the objective of Gradient Descent?

Ans: Objective of gradient descent is to achieve the model that minimizes the loss function.

Q.15 What is the criterial to have the minimum of a cost function.

Ans: The function needs to be a differentiable convex function.

Q.16.How SSE(sum of squared errors) is related to the gradient?

Ans: Gradients give the direction of the movement of m and b wrt SSE, m and b are coefficients or variables of the function.

- Finally, the gradient is made up of all the partial derivatives i.e. both partial derivatives of SSE as shown below:

$$\nabla SSE = \begin{bmatrix} \frac{\partial}{\partial m} SSE \\ \frac{\partial}{\partial b} SSE \end{bmatrix} = \begin{bmatrix} -(Y - \overline{Y})X \\ -(Y - \overline{Y}) \end{bmatrix}$$

- gradients give the direction of the movement of m and b w.r.t to SSE.

Q.17 What is learning rate?

Ans: The learning rate (alpha) is the size of the step gradient descent takes all the way until it reaches the global minimum.

Q.18 Write the name of a second order optimization method.

Ans: Newton Raphson

Q.19 What was the expectation from Newton's method initially?

Ans: Newton method is originally intended to find root of an equation.

Q.20. Write the formula of general equation by Newton method.

on $$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}.$$

Q.21 Probabilistic Interpretation of Linear Regression - The errors are distributed in a _____ distribution with mean as _____ and variance as _____.

Ans: Gaussian/normal, mean as zero, variance as sigma 2

Q.22 Write the formula used in -----

a) probability density for the error

This means that the probability density of $\epsilon^{(i)}$) is:

$$P(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right)$$

b) probability density for the output

Q.23 Define likelihood and Maximum Likelihood Estimate.

Ans: Likelihood is a quantitative/statistical measure of how well a model fits a given set of data.

Maximum Likelihood Estimation: The goal of likelihood estimation is to find the parameter values that maximise the likelihood functions. These parameter values are called MLE

The parameter values used to maximise the likelihood functions are called Maximum Likelihood Estimate

Q24. What are the limitations of gradient descent?

Ans:

1. Calculating derivatives for the entire dataset is time consuming
2. Memory required is proportional to the size of the dataset.

SVM

Q.1) What is Support Vector Machine? Write 3 applications of SVM .

Ans: Support Vector Machine is a supervised machine learning algorithm for classification and regression problems.

Three applications are:

- Used for stock market forecasting by various financial institutions
- Used to compare the relative performance of the stocks in the same sector
- Relative comparison of stocks helps manage investment making decisions based on the classifications made by the SVM learning algorithm.

Q.2) Define the terms:

Support vectors, Hyperplane/ decision boundary, Margin, Soft Margin, Hard Margin. (Use diagram)

Support vectors: The data points or vectors that are the closest to the hyperplane are called support vectors.

Hyperplane(decision boundary): SVM finds a decision boundary to seperate different classes by maximizing the margin.

Margin: The distance between either side of the dashed line to the solid line is the margin

Soft Margin: Softmargin is a technique permitted by SVM, when the data is not perfectly seperable or contains outliers. The misclassified dots that can be tolerated are soft margin.

Hard margin: The hard margin hyperplane is a hyperplane that properly separates the data points of different categories without any misclassifications.

Q3) Distinguish : Linear and Non linear SVM

Ans: Linear SVM

- Used for linearly separable data
- If a data set can be classified into two classes by using a single straight line then such data is termed as linearly separable data.
- The classifier used is called as Linear SVM classifier.

Non Linear SVM

- Used for non linearly separable data.
- If a dataset cannot be classified by using a straight line then such data is termed as non-linear data.
- Classifier used is called as non linear svm classifier

Q.4) How to plot decision boundary if the dataset is linearly separable and it has only 2 input features.

If there are two input features, then the hyperplane will be a straight line.

Q.5) How to plot decision boundary if the dataset is linearly separable and it has more than 2 input features.

Ans: A plane

Q.6) How to plot decision boundary of the dataset is NON-linearly separable? Do you think kernel trick can help in this? If yes, how?

By using the polynomial kernel method, because it converts non linearly separable to separable problems so that the hyperplane can be easily find out even if the data points are not linearly separable in the original input space.

Q.7) Write the mathematical equation of hyperplane.

- Any Hyperplane can be written mathematically as followed:

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots + \beta_n * x_n = 0$$

- For a 2-dimensional space, the Hyperplane, which is the line, could be written as:

$$\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 = 0$$

Q.8) What is degree of tolerance? how to use penalty term in SVC(). Explain.

Degree of tolerance is how much tolerance we want to set when finding the decision boundary is an important hyper-parameter for the SVM (both linear and nonlinear solutions).

Q.9) Write the Python code for SVM implementation. Hyper parameter tuning, and Coefficient.

Q.10) Define kernel method. What are the types of Kernel function - briefly explain

Q.11) Elaborate- Kernel properties

Q.12) What is Kernel Density Estimate?

Q.13) What is Gaussian Kernel function?

Q.14) Define bandwidth optimization. How small or low bandwidth in KDE curve impacts the prediction/ classification by SVM?

Q.15) How to use Kernel trick for - supervised and unsupervised ML?