1. **What is linear regression?**

Linear Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data.

2. **What is Cost Function?**

Cost Function is a mathematical function that measures the error or difference between predicted values and actual values in a machine learning model.

3. **What is Gradient Descents?**

Gradient Descent is an optimization algorithm used to minimize the cost function by iteratively adjusting the model parameters in the direction of the steepest descent (negative gradient).

In the cost function for linear regression:

$$J(\theta) = \frac{1}{2m} \sum (h_\theta(x^{(i)}) - y^{(i)})^2$$

4. **What is Convergence Theorem?**

The Convergence Theorem in the context of gradient descent refers to the idea that under certain conditions, the gradient descent algorithm will converge to the global minimum of the cost function.

5. **What is Global Minima?**

Global Minima:

The global minimum is the absolute lowest point of the entire function across its entire domain.

- At the global minimum, the function has its smallest possible value.
- There can only be one global minimum for convex functions.

6. **What is Local Minima?**

Local Minima:

A local minimum is a point where the function value is lower than nearby points, but not necessarily the lowest overall.

- There can be multiple local minima.

- Occurs often in non-convex functions (e.g., neural networks).

7. **Does θ (theta) change every step in Gradient Descent?**

**Yes**, in gradient descent, the θ values (parameters) are updated at every step (iteration) to move toward the global minimum of the cost function.

This change is what forms the convex "bowl-shaped" descent graph — as you keep updating θ, your cost function value (J(θ)) keeps decreasing.

8. **How is θ updated?**

Using this formula:

$$\theta := \theta - \alpha \cdot \nabla J(\theta)$$

Where:

- Θ - model parameter (like slope or intercept)
- α: learning rate
- ∇J(θ: gradient of the cost function (i.e., slope of the curve at that point)

9. **What is Learning Rate (α)?**

The learning rate controls how big each step is when updating θ.

- If α is too small → slow learning, takes many steps to reach the minimum
- If α is too large → overshoots the minimum, may never converge
- Ideal α → steady, efficient convergence to the global minimum

10. **What is MSE (Mean Squared Error)?**

MSE is a performance metric that measures the average squared difference between the actual values and the predicted values by a model. It tells you how far off your predictions are, on average.

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Key Points:

- Always non-negative (because of squaring)
- Lower MSE = better model
- Sensitive to outliers (squared errors amplify large differences)

### 11. What is MAE (Mean Absolute Error)?

MAE measures the average absolute difference between the actual values and the predicted values. It tells you, on average, how much your predictions deviate from the true values.

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|h_\theta(x^{(i)}) - y^{(i)}\right|$$

Key Points:

- Uses absolute values instead of squaring
- Less sensitive to outliers than MSE
- MAE is in the same unit as the target variable (unlike MSE)

### 12. What is RMSE (Root Mean Squared Error)?

RMSE is the square root of the Mean Squared Error. It measures the average magnitude of the error between predicted and actual values — in the same units as the target variable.

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2}$$

Key Points:

- RMSE is just the square root of MSE
- More interpretable than MSE since it's in original units
- Still sensitive to outliers (because of squaring before root)