Gradient descent is an optimization algorithm commonly used in machine learning and mathematical optimization. Its primary purpose is to minimize a function by iteratively moving towards the direction of steepest decrease in the function's value. This optimization technique is widely employed in training machine learning models to find the optimal parameters that minimize a cost or loss function.

Key concepts:

1. Objective Function:
In the context of machine learning, the function we want to minimize is often referred to as the "objective function" or "cost function." This function measures the difference between the predicted values of a model and the actual values (labels) for a given set of input data.

2. Parameters:
The parameters of a model are the variables that the algorithm adjusts during the training process to minimize the objective function. These parameters could be the weights in a neural network, coefficients in a linear regression model, or any other variables that affect the predictions.

3. Gradient:
The gradient of a function is a vector that points in the direction of the steepest increase of the function at a particular point. In the context of optimization, the negative gradient points in the direction of the steepest decrease. The gradient is calculated using the partial derivatives of the objective function with respect to each parameter.

4. Learning Rate:
The learning rate is a hyperparameter that determines the size of the step taken during each iteration of the optimization process. A too-small learning rate may cause slow convergence, while a too-large learning rate may cause the algorithm to overshoot the minimum and fail to converge.

Steps of Gradient Descent:

1. **Initialize Parameters:**

Start with initial values for the parameters of the model.

2. **Compute Gradient:**

Calculate the gradient of the objective function with respect to each parameter. This is done by computing the partial derivatives of the objective function.

3. **Update Parameters:**

The update rule is as follows:

New Parameter = Old Parameter − Learning Rate × Gradient

Repeat steps 2 and 3 until convergence or a specified number of iterations is reached.

Types of Gradient Descent:

Batch Gradient Descent:

In each iteration, the entire dataset is used to compute the gradient of the objective function.

Stochastic Gradient Descent (SGD):

In each iteration, a random sample (one or a few data points) is used to compute the gradient.

Mini-Batch Gradient Descent:

A compromise between batch and stochastic gradient descent, where a small batch of data is used in each iteration.