1. Since that the gradient descent from the optimizers , so what are the optimizers?

Optimizers are algorithms used to minimize the loss function during the training process of machine learning models, particularly in deep learning neural networks. The loss function quantifies how well the model is performing relative to the desired outcome. The goal of optimization is to find the set of parameters (weights and biases) that minimize this loss function.

Some common optimizers used in deep learning include:-

Stochastic Gradient Descent (SGD): This optimizer updates the parameters using the gradient of the loss function with respect to a subset of the training data (a mini-batch) at each iteration.

Adam (Adaptive Moment Estimation): Adam combines ideas from RMSProp and Momentum. It adapts the learning rates for each parameter based on estimates of the first and second moments of the gradients.

1. What are the types of gradient descent and summarize each of them?

Batch Gradient Descent (BGD): In BGD, the entire dataset is used to compute the gradient of the loss function with respect to the parameters.

Stochastic Gradient Descent (SGD): SGD updates the parameters using the gradient of the loss function with respect to a single randomly chosen sample (or a small subset) from the training dataset at each iteration.

Mini-batch Gradient Descent: It combines the advantages of both BGD (smooth convergence) and SGD (efficiency with large datasets).

-Where updates are computed using a small random subset (mini-batch) of the training dataset.

1. What is the effect of choosing learning rate in Batch gradient descent?

The learning rate plays a crucial role in determining the convergence speed and stability of the optimization process.

It involves a trade-off between convergence speed, stability, and robustness to noise

The learning rate controls the size of the steps taken towards the minimum of the loss function.

1. What is the goal of using gradient descent with a machine learning model?

To minimize the loss function.

Gradient descent is an optimization algorithm that iteratively updates the parameters of the model in order to minimize this loss function.

1. What is the formula for gradient descent in linear regression?

The formula for the gradient descent update rule for linear regression involves calculating the gradient of the mean squared error (MSE) loss function with respect to each parameter. The mean squared error is commonly used as the loss function for linear regression and is defined as:

Where:

m is the number of training examples,

is the actual output for the ii-th training example,

is the predicted output for the ii-th training example.