Gradient Desent:

Main task to update the parameter of neural network by minimizing the loss function respect to parameters

Optimization alogrithms:

the main task of optimization alogrithms is to reduce the loss by adjusting weights and biases.

Batch Gradient Descent:

theta := theta - n(eta) 1/N limt (sum(i=1 to N) delta L)

Advantages:

stable convergence

Disadvantages:

Computationally expensive

SGD:

theta := theta - n(eta) delta L

Adantages:

1. Faster convergence
2. Due to randomness can escape local minima.

Disadvantages:

1. as this is updating single single parameters that's why found high variences is updating parameters nad creating high fluctuating .

Mini-Batch Gradient Descent:

theta := theta - n(eta) (sum(limt i=1 to M) delta L)

Adantage:

it reduces variance of the paramter as it is wokring like a batch.

Advance Optimization Alogorithms:

to improve basic gradient descent problem, adding momentum to improve learning rate and accelerating convergence.

Momentum:

by adding previous value to current value make a stable and smoothing fluctuating and accelerate the convergence and prevent noise.

first means at initial stages all are 0,

like velocity v = 0

parameter, theta = 0

update ,

velocity, V = lemda v + n(eta) delta L here add lemda as a momentum

parameters update, theta = old theta - V(update velocity)