Page No.
Date

Gradient descent

Optimization algorithm to minimize a cost function
Goal is to find the optimal set of parameters the minimizes cost function
Primary idea is the to move in the direction of steepest descent of a function to reach local minimum

Learning rate :- It is a hyperparameter that determines the size of the steps taken

$$x\_new = x\_old + \alpha * f'(x)$$

$$f(x) = x^2 - 2x + 2$$

$$f'(x) = 2x - 2$$

for suppose $x = 3$ & $\alpha = 0.2$

$$\therefore x\_new = x\_old + 0.2 \times f'(3)$$

$$= 3 + 0.2 \times (6 - 2)$$

$$= 3 + 0.2 \times 4$$

$$x\_new = 3 - 0.8 = 2.2$$

Page No.
Date

Now $x\_old = 2.2$

$\therefore x\_new = 2.2 - 0.2 \times f'(2.2)$

$= 2.2 - 0.2 \times (2 \times 2.2 - 2)$

$= 2.2 - 0.2 \times 2.4$

$= 2.2 - 0.48$

$= 1.72$

Repeat the steps until you reach the minimum

Batch vs Mini batch

Batch gradient descent	Mini-batch gradient descent
The entire training dataset is used in each iteration to compute the gradient	It divides the dataset into smaller batches & compute the gradient for each batch.
Accurate gradient	can be prarallelized
Computationally intensive	faster convergence
Stable updates parameters are more stable	

Page No.
Date

Slow convergence Noisy updates.

Stochastic gradient Descent

Variant of gradient descent which processes a single randomly selected training example instead of processing the entire dataset in each iteration