

Definitions

Epoch

An **epoch** is **one complete pass** through the **entire training dataset**.

Iteration

An **iteration** is **one update step** of the model's parameters (weights).

- In **Batch GD**: 1 iteration = 1 epoch
- In **SGD**: 1 epoch = n iterations (where n is number of training examples)
- In **Mini-Batch SGD**:

Iterations per epoch = $\frac{\text{Number of training examples}}{\text{Batch size}}$

Gradient Descent (GD)

Uses the **entire training dataset** to compute the gradient and update weights **once per epoch**.

- **Formula:**

$$w = w - \eta \cdot \nabla L(\text{all data})$$

- **Iterations per epoch:**

$$1 \text{ iteration} = 1 \text{ epoch}$$

Stochastic Gradient Descent (SGD)

Updates weights **after every single data point** (sample).

- **Formula:**

$$w = w - \eta \cdot \nabla L(x_i, y_i)$$

- **Iterations per epoch:**

Number of samples = $N \Rightarrow 1 \text{ epoch} = N \text{ iterations}$ $\text{Number of samples} = N \Rightarrow \text{1 epoch} = N \text{ iterations}$

Mini-Batch Stochastic Gradient Descent

Updates weights after a **mini-batch** of data (e.g., 32, 64, 128 samples).

- **Formula:**

$$w = w - \eta \cdot \nabla L(\text{mini-batch}) \quad w = w - \eta \cdot \nabla L(\text{mini-batch})$$

- **Iterations per epoch:**

N/B (where N = number of samples, B = batch size) $\frac{N}{B} \quad \text{(where } N = \text{number of samples}, B = \text{batch size})$

Example from Image:

- Total samples = 10,000
- Mini-batch size = 1000

$$\text{Iterations per epoch} = \frac{10,000}{1000} = 10 \quad \text{Iterations per epoch} = \frac{10,000}{1000} = 10$$

So:

- 1 epoch = 10 iterations
- Each iteration updates weights using 1,000 records