

Training Loops

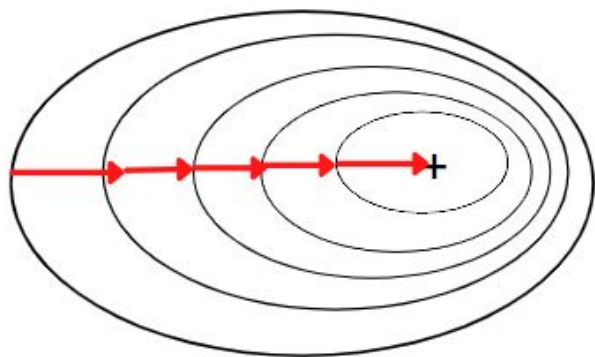
Epoch: A single pass through all of the examples m in the training dataset (a training loop consists of multiple epochs)

Iteration: A single forward and backward pass + weight updates of backpropagation algo (an epoch can consist of multiple iterations)

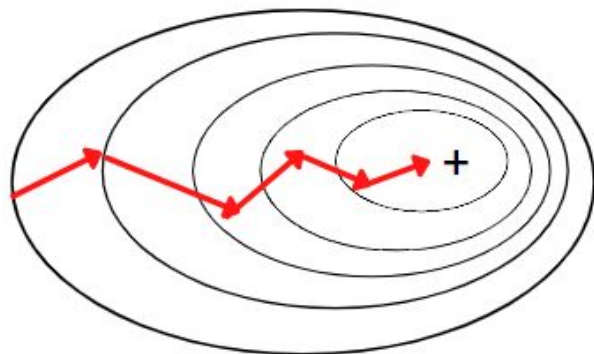
Stochastic vs Batch vs Mini-batch Gradient Descent

- Stochastic: A training dataset of m examples would take m iterations to complete a single epoch, and would involve m weight updates for $w_{i,k}$
- Batch: $\Delta w_{i,k}$ is the sum of error gradient for weight $w_{i,k}$ for all m in one epoch. A single iteration is performed in each epoch, therefore a single weight update per epoch.
- Mini-batch: $\Delta w_{i,k}$ is the sum error gradient for weight $w_{i,k}$ for one mini-batch (subset of m). A training dataset of m examples would take $m / \text{batch_size}$ iterations to complete a single epoch, and involve $m / \text{batch_size}$ weight $w_{i,k}$ updates

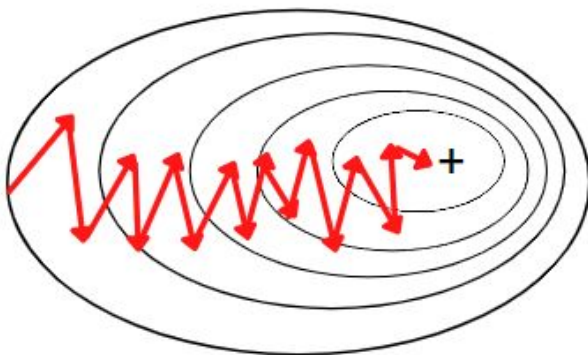
Batch Gradient Descent



Mini-Batch Gradient Descent

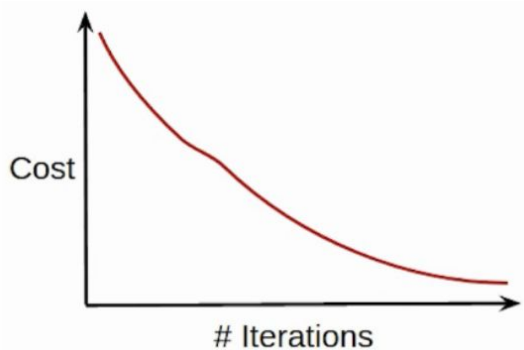


Stochastic Gradient Descent



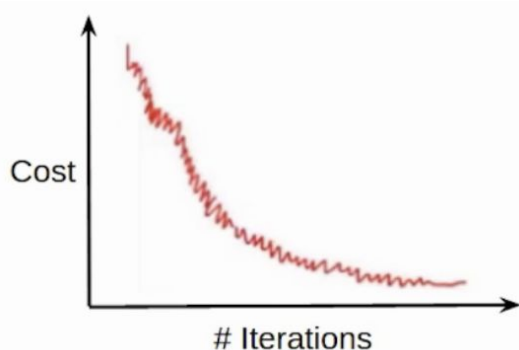
Batch Gradient Descent

- Update using all data
- Smooth cost function
- High computational cost



Stochastic Gradient Descent (SGD)

- Update using single observation
- High variation in cost function
- Low computational cost



Mini-Batch Gradient Descent

- Update using subset of data
- Smoother cost function compared to SGD
- Balance between Batch / SGD

