Reducing Loss: Check Your Understanding

Estimated Time: 2 minutes

Check Your Understanding: Batch Size

Explore the options below.

When performing gradient descent on a large data set, which of the following batch sizes will likely be more efficient?

✓ A small batch or even a batch of one example (SGD).

Amazingly enough, performing gradient descent on a small batch or even a batch of one example is usually more efficient than the full batch. After all, finding the gradient of one example is far cheaper than finding the gradient of millions of examples. To ensure a good representative sample, the algorithm scoops up another random small batch (or batch of one) on every iteration.

Correct answer.

🚫 The full batch.

Computing the gradient from a full batch is inefficient. That is, the gradient can usually be computed far more efficiently (and just as accurately) from a smaller batch than from a vastly bigger full batch.

Try again.

Key Terms

- batch (https://developers.google.com/machine-learning/glossary#batch)
- mini-batch
- batch size (https://developers.google.com/machine-learning/glossary#batch_size)
- stochastic gradient descent