Minibatch Stochastic Gradient Descent
SpeakUp

Join by URL: https://web.speakup.info/room/join/80740

Join by key: web.speakup.info → use key 80740

Join by saved session: web.speakup.info

Join by QR:
Outline

• General neural network process

• (Batch) gradient descent

• Stochastic gradient descent (SGD)

• Minibatch stochastic gradient descent (Minibatch SGD)

• Discuss tradeoffs
Recap: Automatic Differentiation

• Take five minutes to draw
  • Whatever will help you remember (no correct or incorrect drawings)
  • You’ll keep a running drawing log the rest of the semester
General Neural Network Process

1. Prepare your dataset
   • Normalize inputs
   • Create a proxy (smaller) dataset for debugging and testing workflows
   • Split the dataset into training, validation, and evaluation

2. Design a neural network (architecture)

3. Set initial hyperparameters (learning rate, number of epochs, etc.)

4. Train model parameters (includes validation)

5. Evaluate the trained model

6. Deploy the trained model
(Batch) Gradient Descent
Stochastic Gradient Descent
Minibatch Stochastic Gradient Descent
Tradeoffs

• (Batch) gradient descent
  • Fewer updates means slower progress in terms of real time
  • Maximally leverages parallel computations (faster)

• Stochastic gradient descent
  • Many updates can sometimes lead to faster convergence
  • Noisy data means that some updates are bad

• Minibatch stochastic gradient descent
  • Noisy but less than SGD
  • Pretty fast but slower than batch gradient descent
  • Should always be your default choice
  • Add another hyperparameter called `batch_size`
Code Demo

https://github.com/anthonyjclark/cs152sp23/blob/main/Lectures/06-MinibatchSGDDemo.ipynb