PAGE
DATE

(week 2) Optimization Algorithms.

. It enables us to train the NN much faster

ⓐ Mini batch gradient descent ← Batch

. with the implementation of gradient descent on the whole training set what you have to do is to process the entire training set before taking one little step of GD

. we can get a faster algorithm if let GD start make some progress even before finishing processing the entire training set. So we split the training set into smaller sets called Mini batches

$$X_{(n_x, m)} = [\underbrace{x^1 \; x^2 \; x^{1000}}_{X^{\{1\}}} \; | \; \ldots \; x^m] \rightarrow \text{same for } ⓨ$$

ⓑ . Batch gradient descent = GD → when you process your entire training set at the same time

. miniBatch refers to algorithm which process single mini batch at time & a single pass through the training set that is one epoch allows you to take N GD steps

ⓒ stochastic GD , BGD . Mini batch GD

→ exponentially weighted Averages.

. is technique used to smooth out data by giving more weights to recent values while considering past values. It is used in optimization algorithms to smooth gradient updates & accelerate convergence

PAGE
DATE

→ gradient descent with momentum (3)
. The basic idea is to compute the weighted avg of gradients & then use that gradient to update the weight insted
. Standard GD relies on the current gradient for updated which lead to slow convergence & oscillations, GD with momentum leverages past gradients to stabilize & accelerate convergence

→ RMsprop (Root mean square prop) (2)
. Is an optimization algorithm that is used for speed up gradient descent

→ Adam optimization algorithm. (1)
. adam takes momentum (to speed up convergence in the right direction) & RMsprop (to adjust learning rate) that lead to robust optimization

→ learning rate decay: one of the ways that might speed up the algorithm learning rate by slowly reduce your learning rate over time, often implemented after set number of epochs or when loss plateaus: which is a situations during training where loss stops decreasing & remain constant over several iterations or epochs