CompSci, December 14, 2022

$$\beta^{(k+1)} = \beta^{(k)} - \gamma^{(k)} g(\beta^{(k)})$$

- constant γ_0
- linear
  $\gamma^{(k)} = (1-\alpha)\gamma_0 + \alpha\gamma_\tau$
  $\alpha = \frac{k}{\tau}$
  γ_τ is constant
  $\gamma_\tau \sim \frac{1}{100}\gamma_0$
- $\gamma^{(k)} = \frac{\gamma_0}{1+k\gamma_\tau}$
- exponential
  $\gamma^{(k)} = \gamma_0 \exp(-k\gamma_\tau)$

- Gradient descent (GD)
- stochastic gradient descent (SGD) with mini-Batches

- Adaptive methods
  - Adagrad (convex $C(\beta)$)
  - RMSprop

– ADAM

ADAgrad

Algorithm:

require: learning rate γ_0

initial guess $\beta^{(0)}$

constant (small) δ

for numerical

stability.

while stopping criterion not met

- minibatches if SGD.
- compute gradient g
- accumulate squared gradient

$r_{i+1} = r_i + g \odot g$

$y \odot x$

$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \odot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= \begin{bmatrix} y_1 x_1 \\ y_2 x_2 \end{bmatrix}$

- compute

$\frac{\gamma_0}{\delta + \sqrt{r}} \cdot g$

– adaptive learning rate

$\cancel{y \odot x}$

$y * x$

- update

$$\beta_{i+1} = \beta_i - \frac{\gamma_0}{\delta + \sqrt{r}} g$$

RMSprop

Require :- global learning rate γ, decay rate ρ

- initial β_0
- small constant δ
- $r = 0$

while stop criterion not met

- SGD minibatch + epochs
- compute gradients
- accumulate squared gradient

$$r \leftarrow \rho r + (1-\rho) g \odot g$$

- compute parameter

$$\Delta\beta = \frac{-\gamma}{\delta + \sqrt{r}} \odot g$$

- apply update

$$\beta \leftarrow \beta + \Delta\beta$$

end while

<u>ADAM</u> : adaptive momentum

require : - initial learning rate γ_0

- momentum rates ρ_1 and ρ_2

  $\rho_1 \approx 0.9$, $\rho_2 \approx 0.999$

- initial β_0

- s_1 and $r = 0$

while stop criterion not met

- SGD sample minibatches + epochs

- compute gradients

- $t \leftarrow t+1$

- update first momentum

  $$s \leftarrow \rho_1 s + (1-\rho_1) \cdot g$$

- update second momentum

$$r \leftarrow \rho_2 r + (1-\rho_2) g \odot g$$

- correct bias in first momentum

$$s \leftarrow \frac{s}{1-\rho_1^t}$$

- correct bias in 2nd momentum

$$r \leftarrow \frac{r}{1-\rho_2^t}$$

- compute update

$$\Delta\beta = -\frac{\gamma_0 \cdot s}{\sqrt{r} + \delta}$$

- update β

$$\beta \leftarrow \beta + \Delta\beta$$

end while.