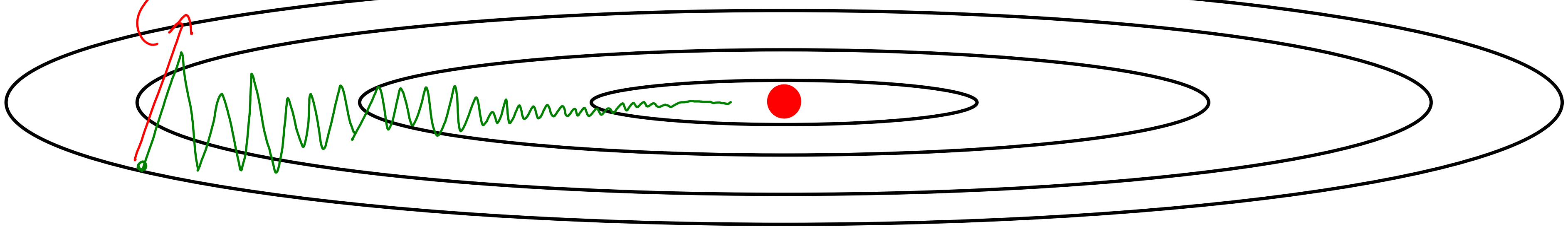
deeplearning.ai

Optimization Algorithms

Gradient descent with momentum

Gradient descent example

Dont want to overshoot to higher cost with Large α



currently there are 2 problems
1. up & down oscillation slows down gradient descent
2. we want to not be limited by learning rate, Right now If we $\uparrow \alpha$ we overshoot & $\uparrow J$

- Also, we want to control α along axis
  - $\Rightarrow$ moving left to Right α should be large
  - $\Rightarrow$ moving top to bottom, α = small

Momentum Gradient descent
↳ works for batch & minibatch

on minibatch "t" {

compute dw, db on minibatch t

Moving Avg formula

$$V_{dw} = \beta \cdot V_{dw} + (1-\beta) \cdot dw$$

$$V_{db} = \beta V_{db} + (1-\beta) db$$

$$w = w - \alpha \cdot V_{dw}$$

$$b = b - \alpha \cdot V_{db}$$

}

why will this $\downarrow$ oscillations in vertical direction & $\uparrow$ movement in horizontal direction

Think about the current derivatives

Along the

- vertical direction
  $\uparrow + \downarrow + \uparrow + \downarrow \ldots = $ Avg is 0
  The V_{dw}, V_{db} is calculating the moving Avg $\Rightarrow$ $\boxed{\text{Moving Avg (vertical)} = 0}$

- Horizontal direction
  $\rightarrow + \rightarrow + \rightarrow + \rightarrow$ = Avg is +ve
  $\Rightarrow$ V_{dw}, V_{db} $\boxed{\text{Moving Avg (Horizontal)} \xrightarrow{+ve}}$

deeplearning.ai

Andrew Ng

Implementation details

same dim as w → $V_{dw} = 0$, $V_{db} = 0$ ← Same dim as b

On iteration t:

Compute dW, db on the current mini-batch

$$v_{dW} = \beta v_{dW} + (1-\beta)dW$$

$$v_{db} = \beta v_{db} + (1-\beta)db$$

$$W = W - \alpha v_{dW}, \;\; b = b - \alpha v_{db}$$

Sometimes, you'll Also see

$$\left.\begin{aligned} V_{dw} &= \beta V_{dw} + dw \\ V_{db} &= \beta V_{db} + db \end{aligned}\right]$$

This is the same as the original its just that you'll have to Adjust α by $\alpha \cdot \frac{1}{(1-\beta)}$

Hyperparameters: α, β

Learning Rate

Generally ? $\beta = 0.9$ works well

takes the last $\frac{1}{1-\beta} = 10$ data points

Generally bias correction is ignored, After 1st 10 iterations, formula corrects itself (weather example)

Andrew Ng