deeplearning.ai

Optimization Algorithms

Mini-batch gradient descent

1

Batch vs. mini-batch gradient descent

Vectorization allows you to efficiently compute on m examples.

Andrew Ng

2

Mini-batch gradient descent

Andrew Ng

3

Optimization Algorithms

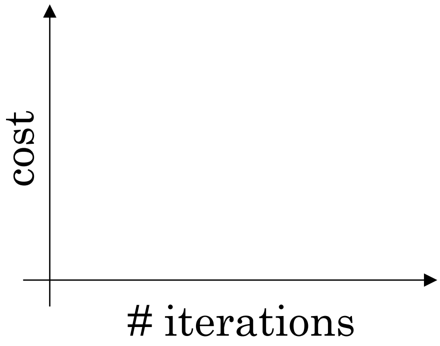
Understanding mini-batch gradient descent

deeplearning.ai

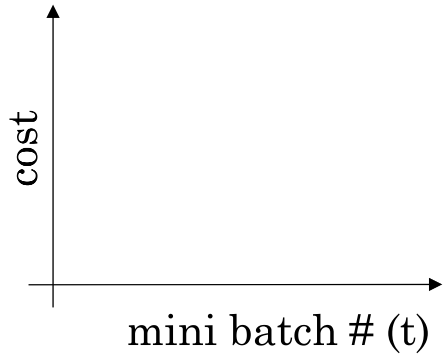
4

Training with mini batch gradient descent

Batch gradient descent



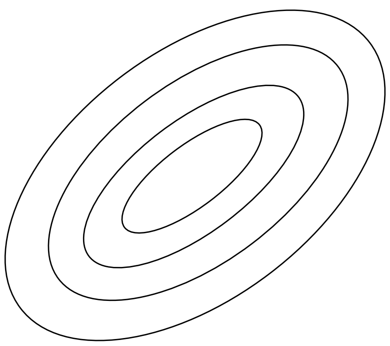
Mini-batch gradient descent



Andrew Ng

5

Choosing your mini-batch size



Andrew Ng

6

Choosing your mini-batch size

Andrew Ng

7

deeplearning.ai

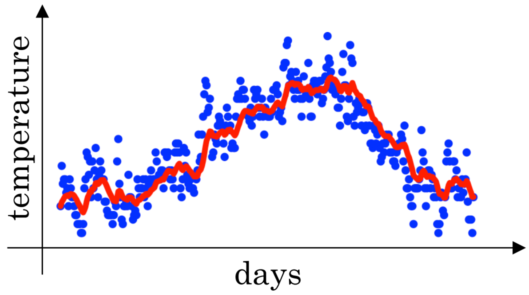
Optimization Algorithms

Exponentially weighted averages

8

Temperature in London

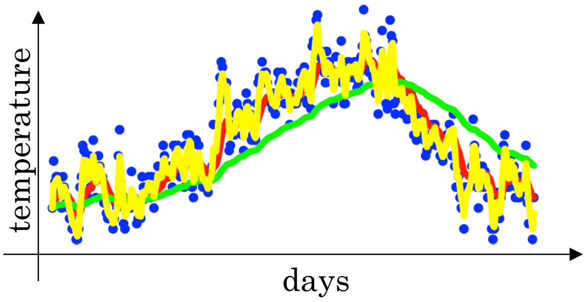
$\theta_1 = 40°F$
$\theta_2 = 49°F$
$\theta_3 = 45°F$
$\vdots$
$\theta_{180} = 60°F$
$\theta_{181} = 56°F$
$\vdots$



Andrew Ng

9

Exponentially weighted averages



Andrew Ng

10

deeplearning.ai

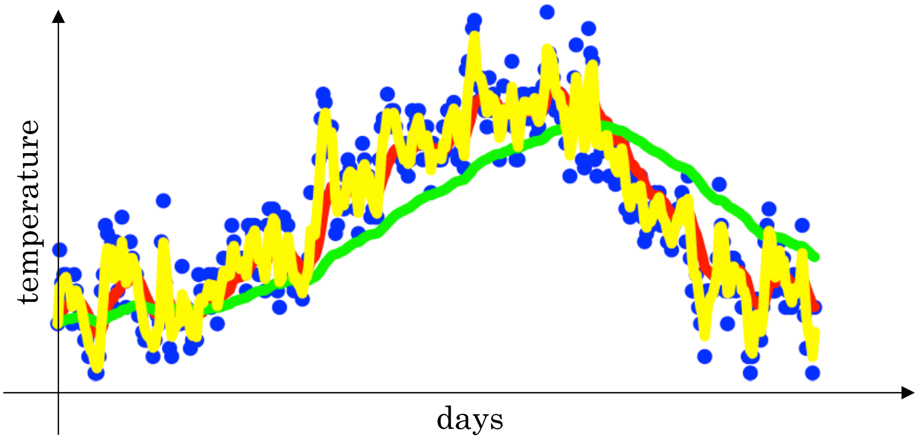
Optimization Algorithms

Understanding exponentially weighted averages

11

Exponentially weighted averages

$$v_t = \beta v_{t-1} + (1-\beta)\theta_t$$



Andrew Ng

12

Exponentially weighted averages

$v_t = \beta v_{t-1} + (1-\beta)\theta_t$

$v_{100} = 0.9v_{99} + 0.1\theta_{100}$

$v_{99} = 0.9v_{98} + 0.1\theta_{99}$

$v_{98} = 0.9v_{97} + 0.1\theta_{98}$

…

Andrew Ng

13

Implementing exponentially weighted averages

$v_0 = 0$

$v_1 = \beta v_0 + (1-\beta)\,\theta_1$

$v_2 = \beta v_1 + (1-\beta)\,\theta_2$

$v_3 = \beta v_2 + (1-\beta)\,\theta_3$

…

Andrew Ng

14

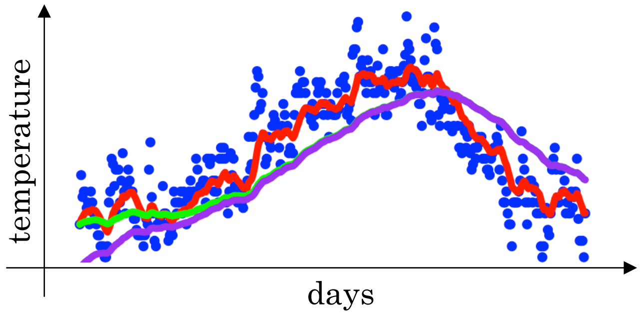
deeplearning.ai

Optimization Algorithms

Bias correction in exponentially weighted average

15

Bias correction



$$v_t = \beta v_{t-1} + (1-\beta)\theta_t$$

Andrew Ng

16

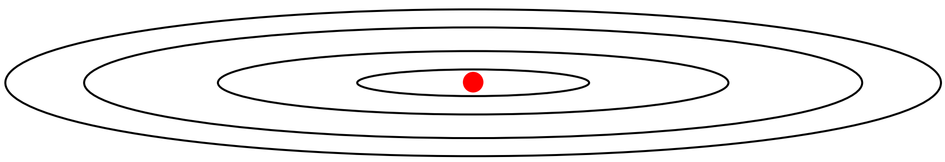
deeplearning.ai

Optimization Algorithms

Gradient descent with momentum

17

Gradient descent example



Andrew Ng

18

Implementation details

On iteration t:

Compute dW, db on the current mini-batch

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

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

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

Hyperparameters: α, β $\beta = 0.9$

Andrew Ng

19

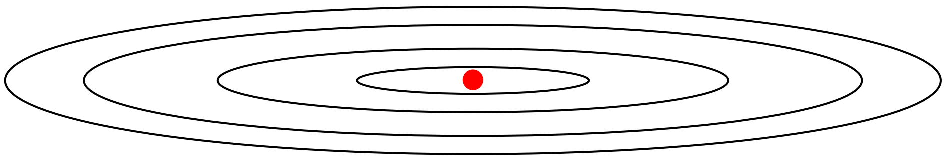
deeplearning.ai

Optimization Algorithms

RMSprop

20

RMSprop



Andrew Ng

21

Optimization Algorithms

Adam optimization algorithm

deeplearning.ai

22

Adam optimization algorithm

Andrew Ng

23

Hyperparameters choice:

Adam Coates

Andrew Ng

24

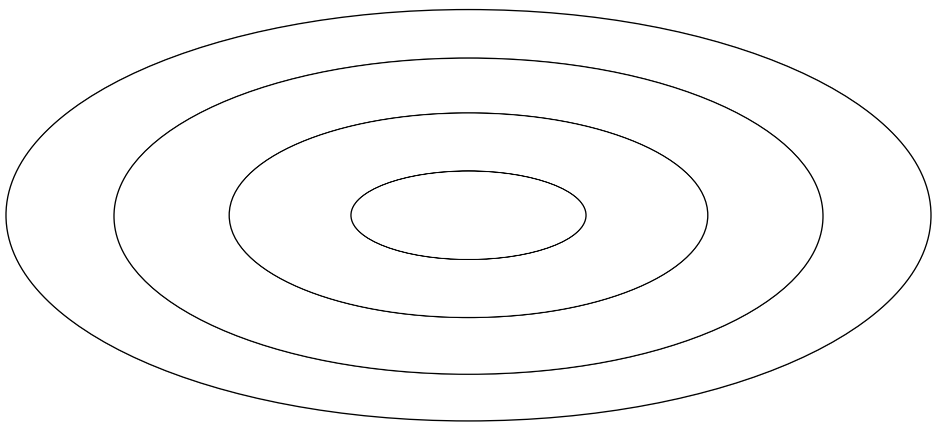
Optimization Algorithms

deeplearning.ai

Learning rate decay

25

Learning rate decay



Andrew Ng

26

Learning rate decay

Andrew Ng

27

Other learning rate decay methods

Andrew Ng

28

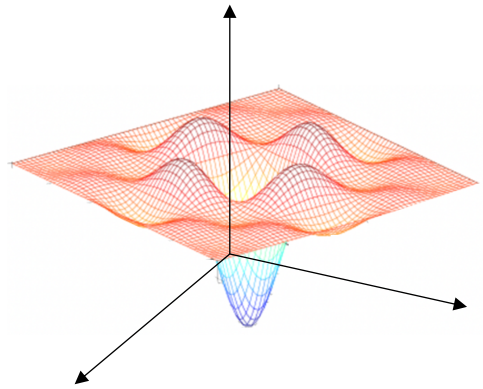
deeplearning.ai

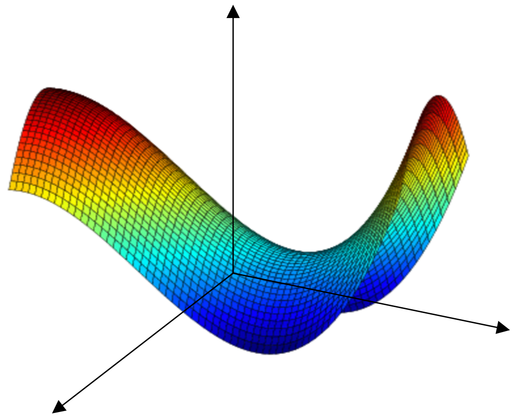
Optimization Algorithms

The problem of local optima

29

Local optima in neural networks

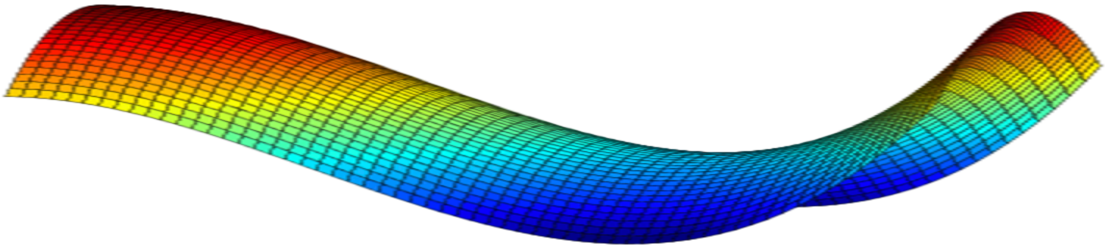




Andrew Ng

30

Problem of plateaus



- Unlikely to get stuck in a bad local optima
- Plateaus can make learning slow

Andrew Ng

31