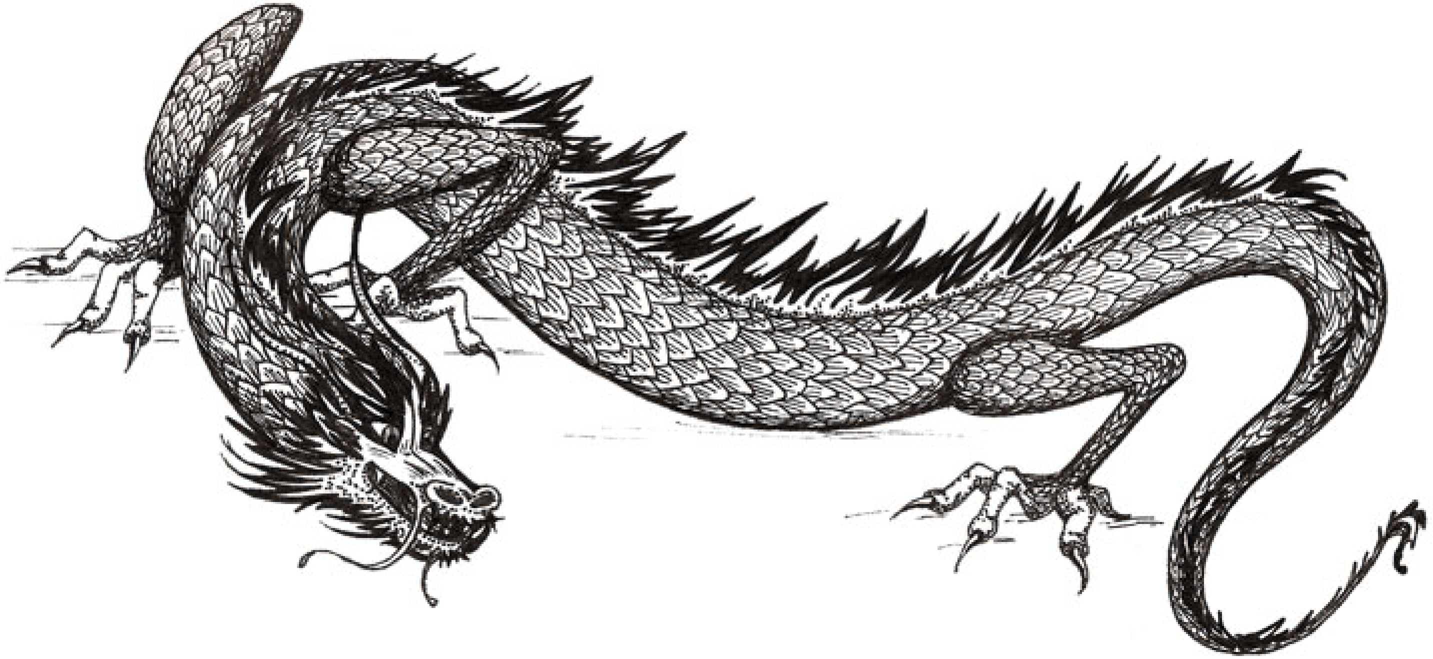


动量与学习率

主讲：龙良曲

Outline

- momentum
- learning rate decay

Momentum

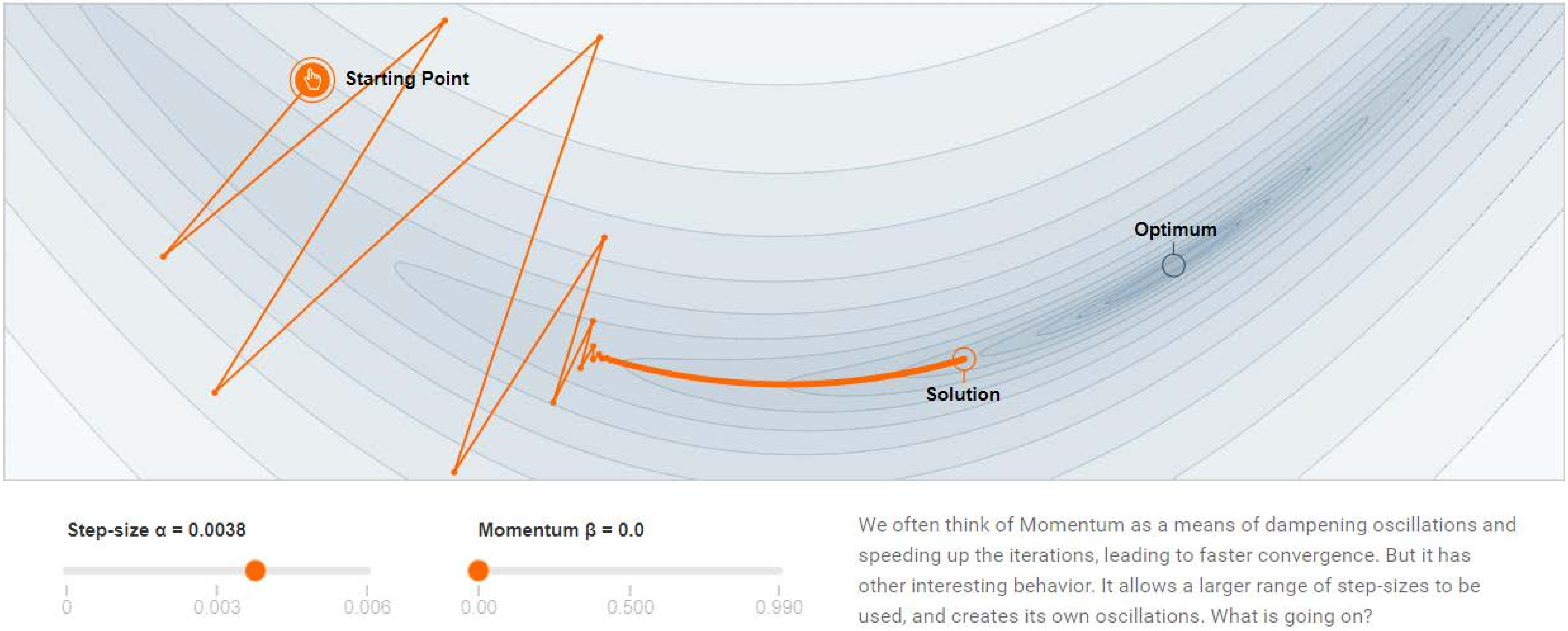
$$w^{k+1} = w^k - \alpha \nabla f(w^k).$$

$$z^{k+1} = \beta z^k + \nabla f(w^k)$$

$$w^{k+1} = w^k - \alpha z^{k+1}$$

https://distill.pub/2017/momentum/

No momentum

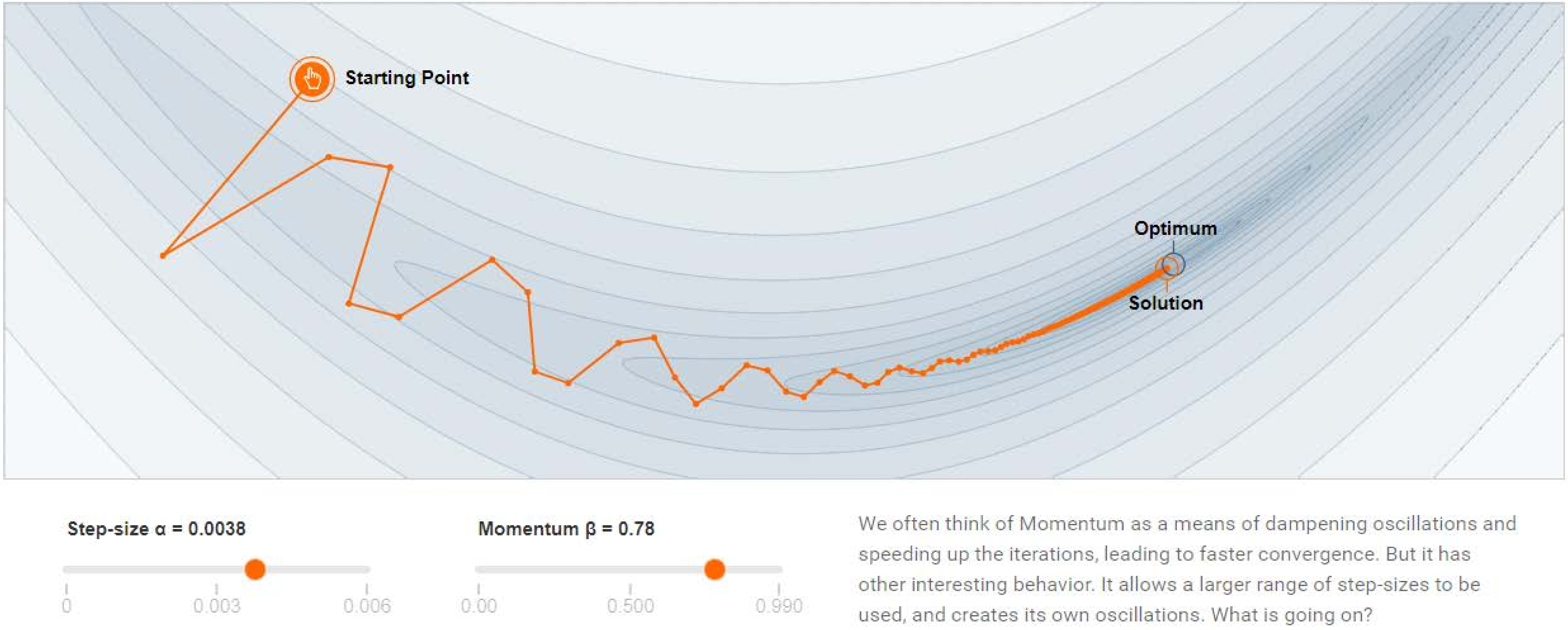


Step-size α = 0.0038

Momentum β = 0.0

We often think of Momentum as a means of dampening oscillations and speeding up the iterations, leading to faster convergence. But it has other interesting behavior. It allows a larger range of step-sizes to be used, and creates its own oscillations. What is going on?

With appr. momentum

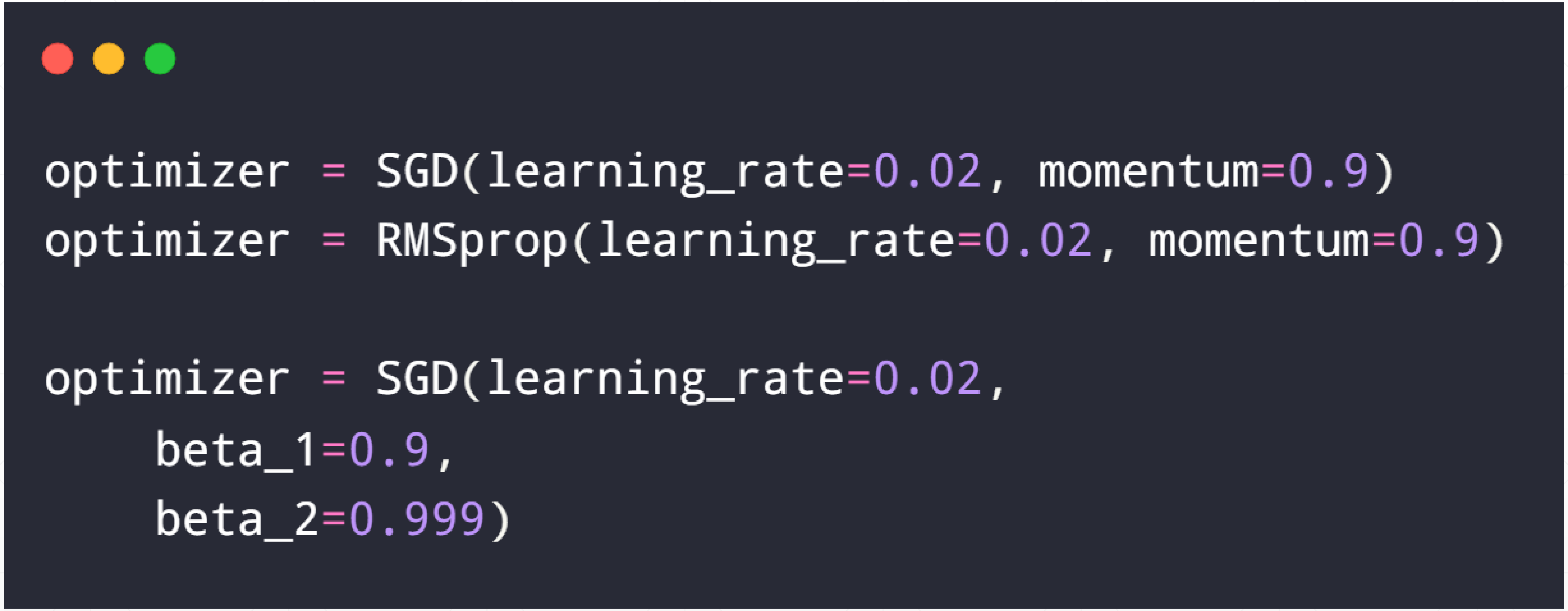


Step-size α = 0.0038

Momentum β = 0.78

We often think of Momentum as a means of dampening oscillations and speeding up the iterations, leading to faster convergence. But it has other interesting behavior. It allows a larger range of step-sizes to be used, and creates its own oscillations. What is going on?

Momentum

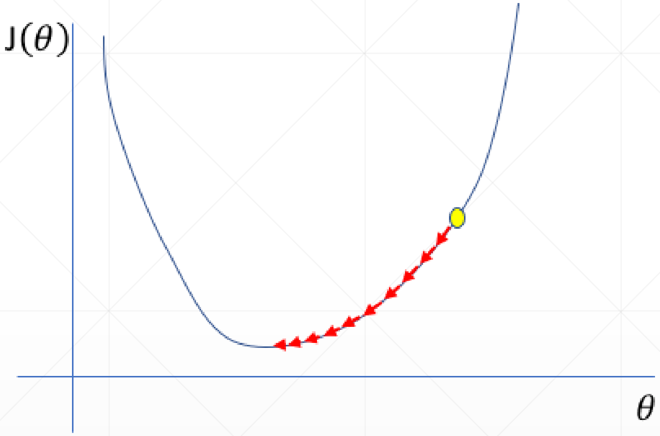


```
optimizer = SGD(learning_rate=0.02, momentum=0.9)
optimizer = RMSprop(learning_rate=0.02, momentum=0.9)

optimizer = SGD(learning_rate=0.02,
    beta_1=0.9,
    beta_2=0.999)
```

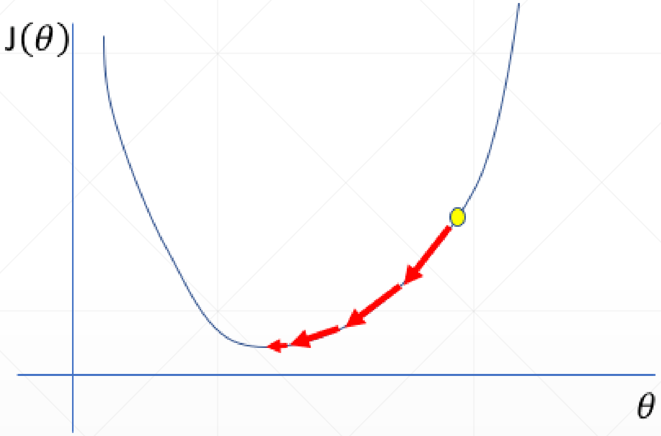
Learning rate tunning

Too low



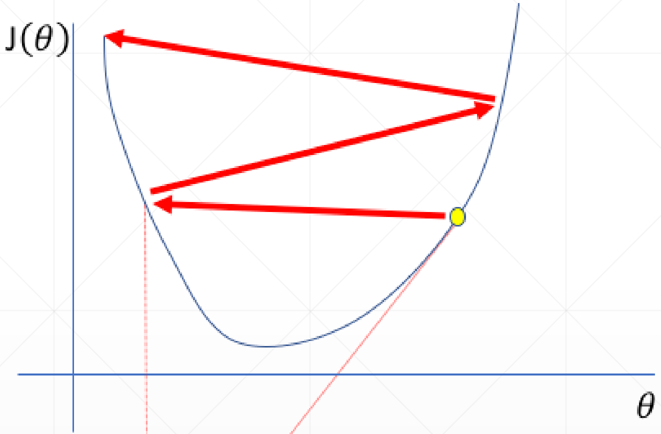
A small learning rate requires many updates before reaching the minimum point

Just right

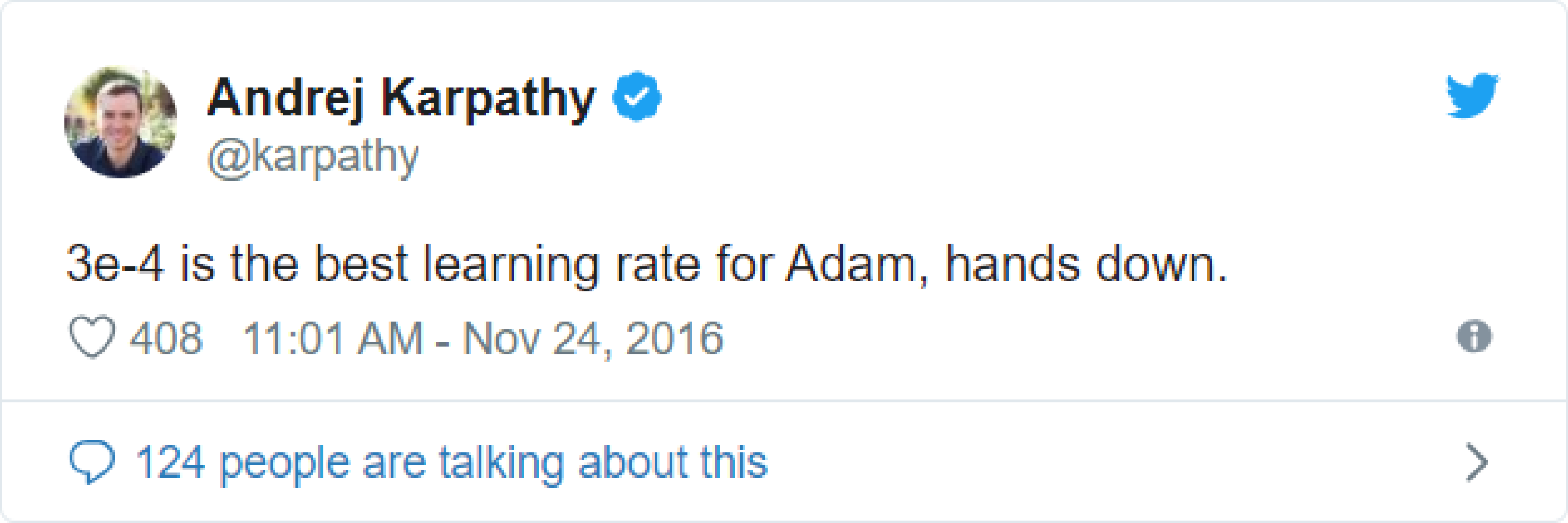


The optimal learning rate swiftly reaches the minimum point

Too high



Too large of a learning rate causes drastic updates which lead to divergent behaviors



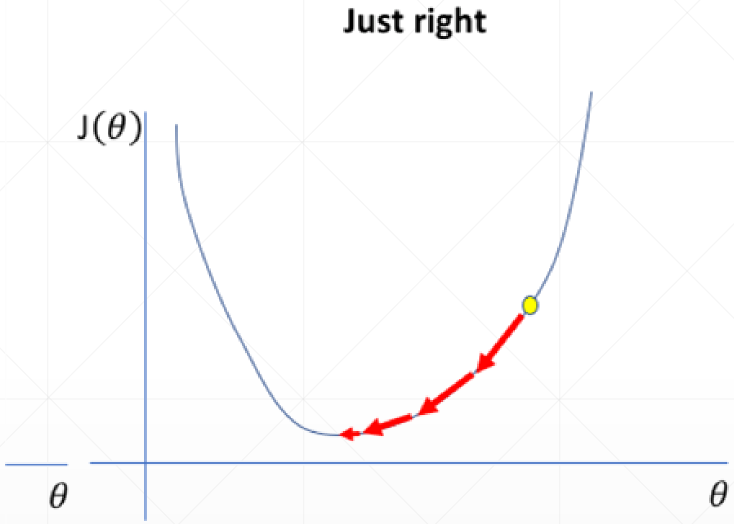
Andrej Karpathy
@karpathy

3e-4 is the best learning rate for Adam, hands down.

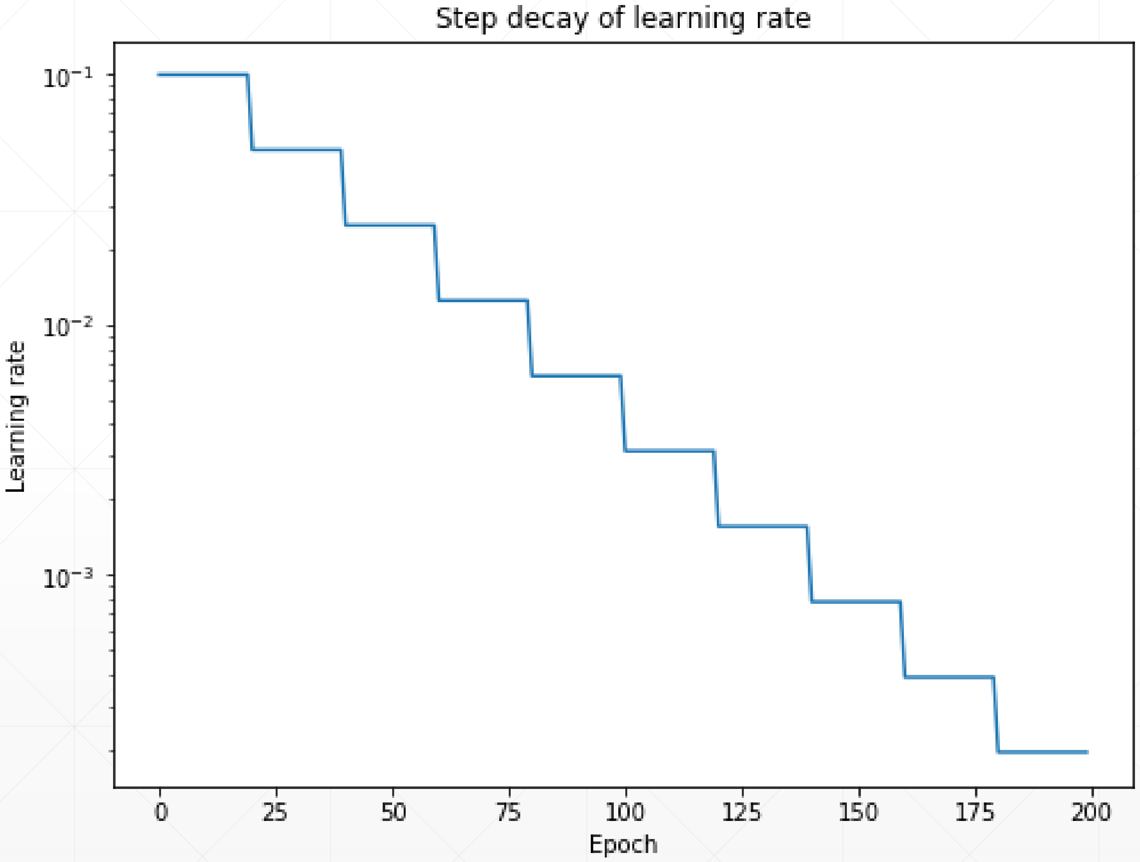
♡ 408 11:01 AM - Nov 24, 2016

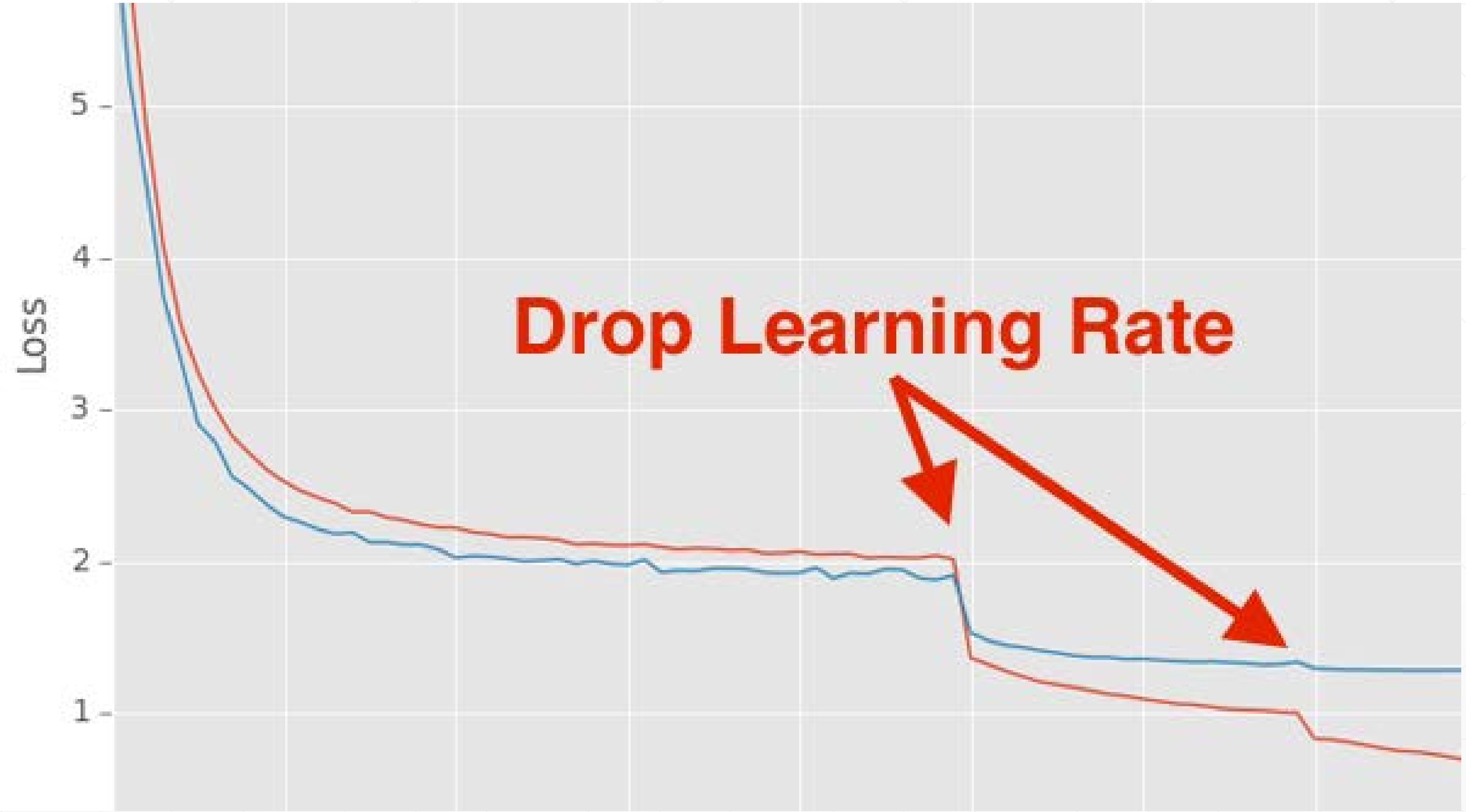
124 people are talking about this

Learning rate decay



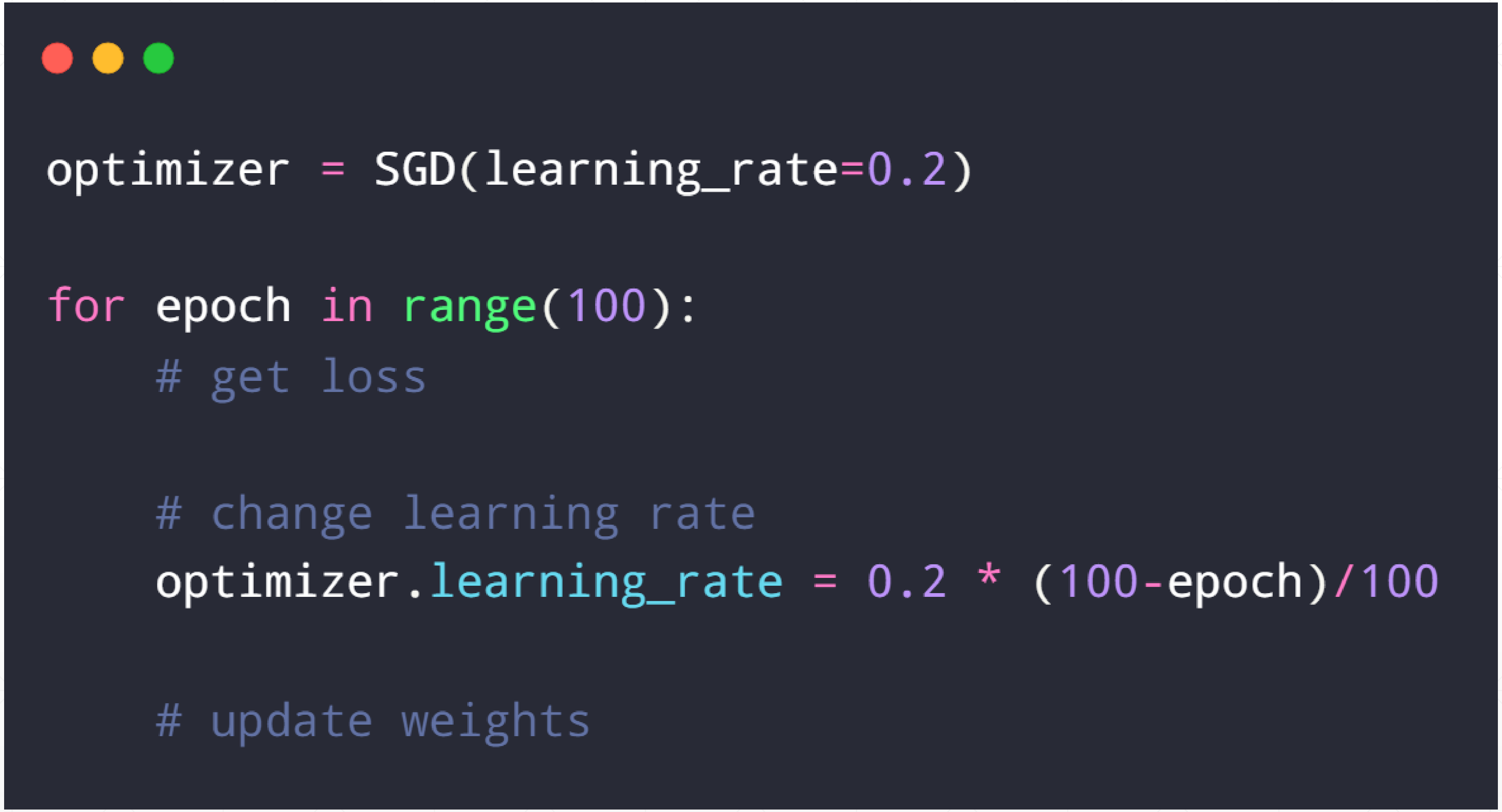
The optimal learning rate swiftly reaches the minimum point





Drop Learning Rate

Adaptive learning rate



```
optimizer = SGD(learning_rate=0.2)

for epoch in range(100):
    # get loss

    # change learning rate
    optimizer.learning_rate = 0.2 * (100-epoch)/100

    # update weights
```

下一课时

Early Stopping, Dropout

Thank You.