

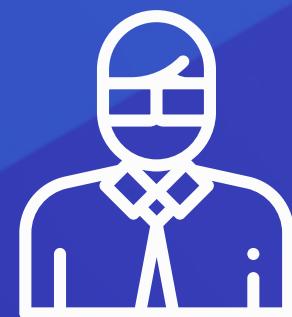


Day 79

初探深度學習使用 Keras

訓練神經網路的細節與技巧

Learning rate effect



游為翔

出題教練

知識地圖 深度學習訓練技巧

Learning Rate

深度神經網路

Supervised Learning Deep Neural Network (DNN)

簡介 Introduction

套件介紹 Tools: Keras

組成概念 Concept

訓練技巧 Training Skill

應用案例 Application

卷積神經網路

Convolutional Neural Network (CNN)

簡介 introduction

套件練習 Practice with Keras

訓練技巧 Training Skill

電腦視覺 Computer Vision

深度學習訓練技巧

Training Skill of DNN

應注意的關鍵

防止過擬合 (Overfitting)

超參數 (Hyper-parameters)

學習率 (Learning Rate) 調整

相關訓練技巧

正規化
Regularization

批次標準化
Batch Normalization

回呼
Callback

隨機移除
Drop out

客製化損失函數
Customized Loss Function

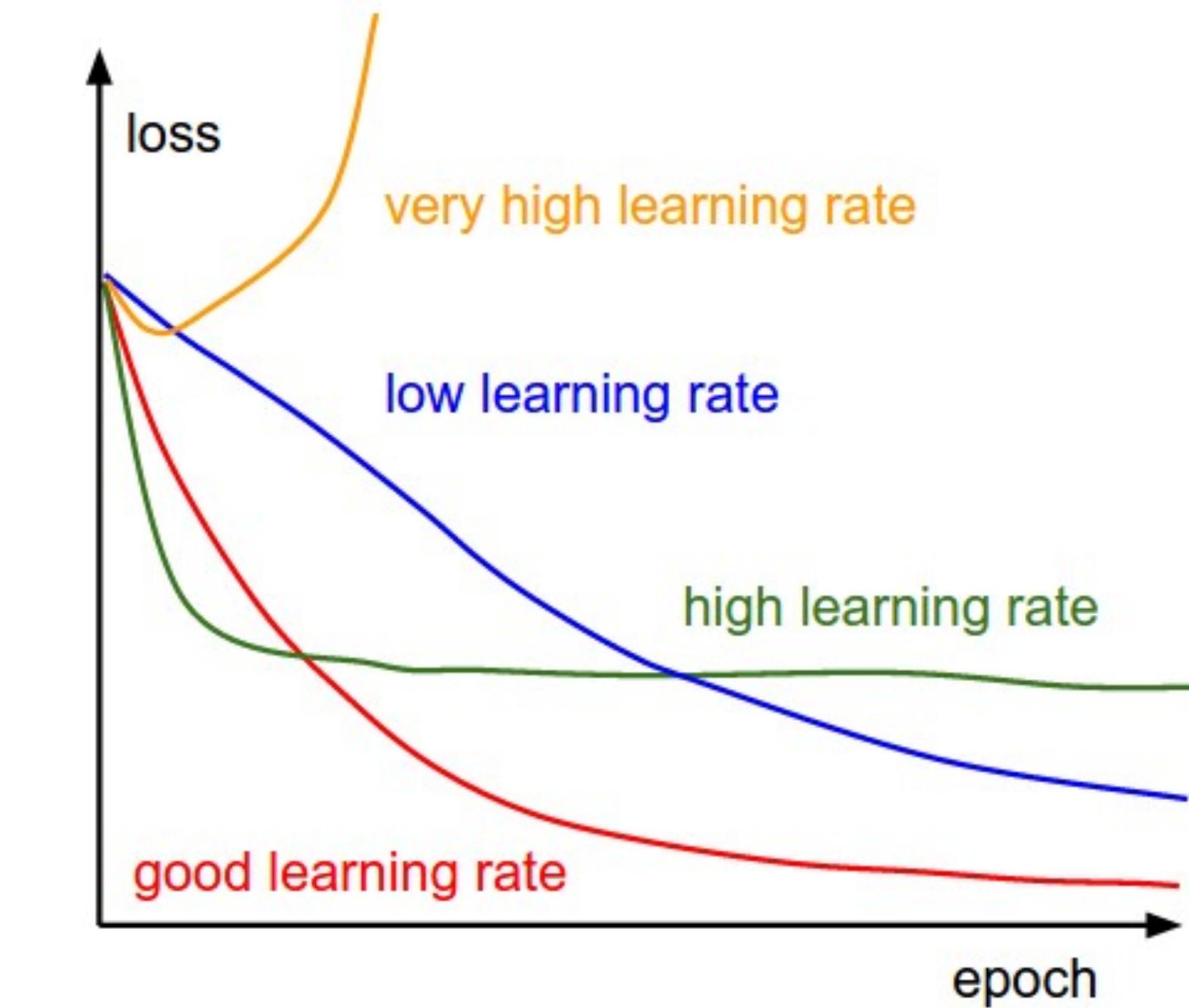
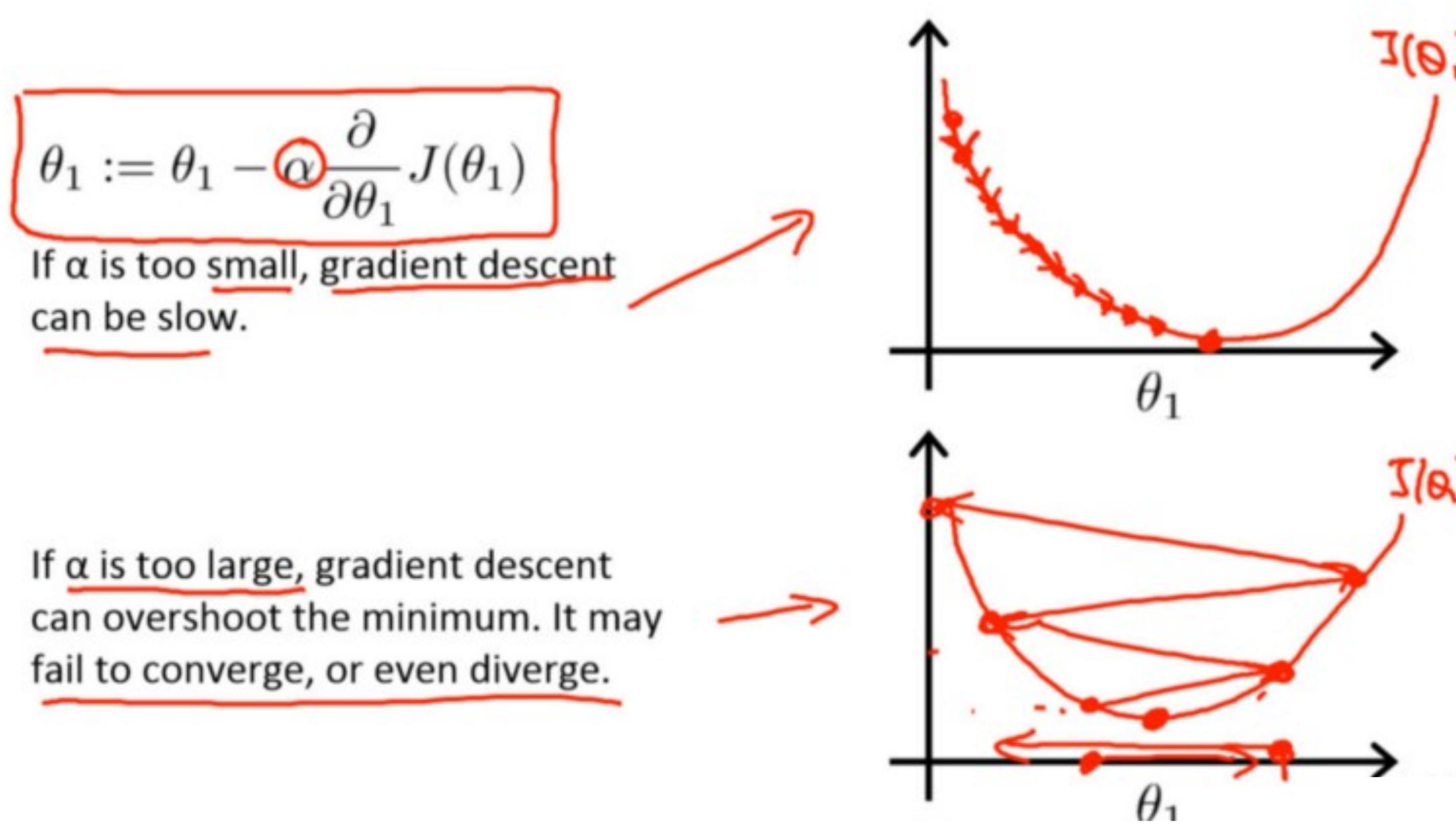
提前終止
Early Stopping

本日知識點目標

- 了解 Learning Rate 對訓練的影響
- 了解各優化器內，不同的參數對訓練的影響

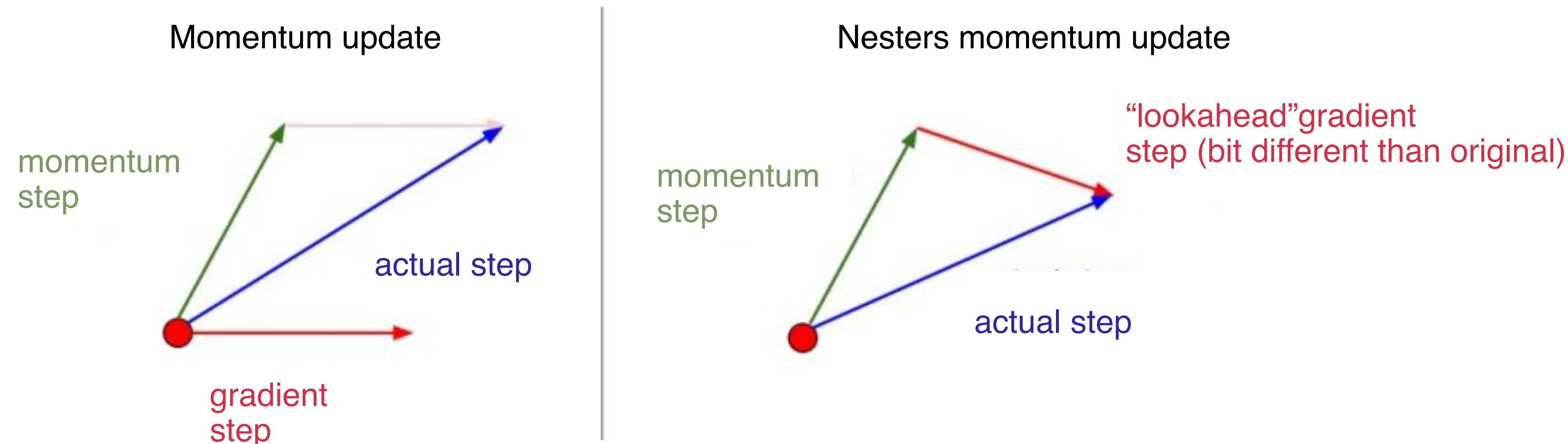
Learning Rate Effect

如果 Learning rate (LR, alpha) 太大，將會導致每步更新時，無法在陡峭的損失山谷中，順利的往下滑動；但若太小，則要滑到谷底的時間過於冗長，且若遇到平原區則無法找到正確的方向。



Options in SGD optimizer

- Momentum : 動量 – 在更新方向以外，加上一個固定向量，使得真實移動方向會介於算出來的 gradient step 與 momentum 間。
 - Actual step = momentum step + gradient step
- Nesterov Momentum : 拔草測風向
 - 將 momentum 納入 gradient 的計算
 - Gradient step computation is based on $x + \text{momentum}$



重要知識點複習

- 學習率對訓練造成的影响
 - 學習率過大：每次模型參數改變過大，無法有效收斂到更低的損失平面
 - 學習率過小：每次參數的改變量小，導致
 - 損失改變的幅度小
 - 平原區域無法找到正確的方向
- 在 SGD 中的動量方法
 - 在損失方向上，加上一定比率的動量協助擺脫平原或是小山谷



延伸 閱讀

除了每日知識點的基礎之外，推薦的延伸閱讀能補足學員們對該知識點的了解程度，建議您解完每日題目後，若有
多餘時間，可再補充延伸閱讀文章內容。

推薦延伸閱讀

知乎 - 深度學習超參數 Learning rate 與 Momentum 理解

- Learning rate : 每次修正的幅度，太大則無法收斂，太小則修正過慢
- Weight decay : 增加正則用以避免 overfitting
- Momentum : 在修正方向上，增加動量，如牛頓定律一樣，增加動量有機會讓卡在局部最小值的狀態跳離
- Learning rate decay : 讓Learning rate 可以隨訓練進行慢慢減小，讓收斂狀

參考連結：

1.[Estimating an Optimal Learning Rate For a Deep Neural Network](#)

2.[cs231n: learning and evaluation](#)

3.[知乎-深度學習超參數簡單理解>>>>learning rate,weight decay](#)

Estimating an Optimal Learning Rate
For a Deep Neural Network



Pavel Surmenok [Follow](#)
Nov 13, 2017 · 5 min read





解題時間

It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

