

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



出題教練

游為翔



Learning Rate

深度神經網路

Supervised LearningDeep 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

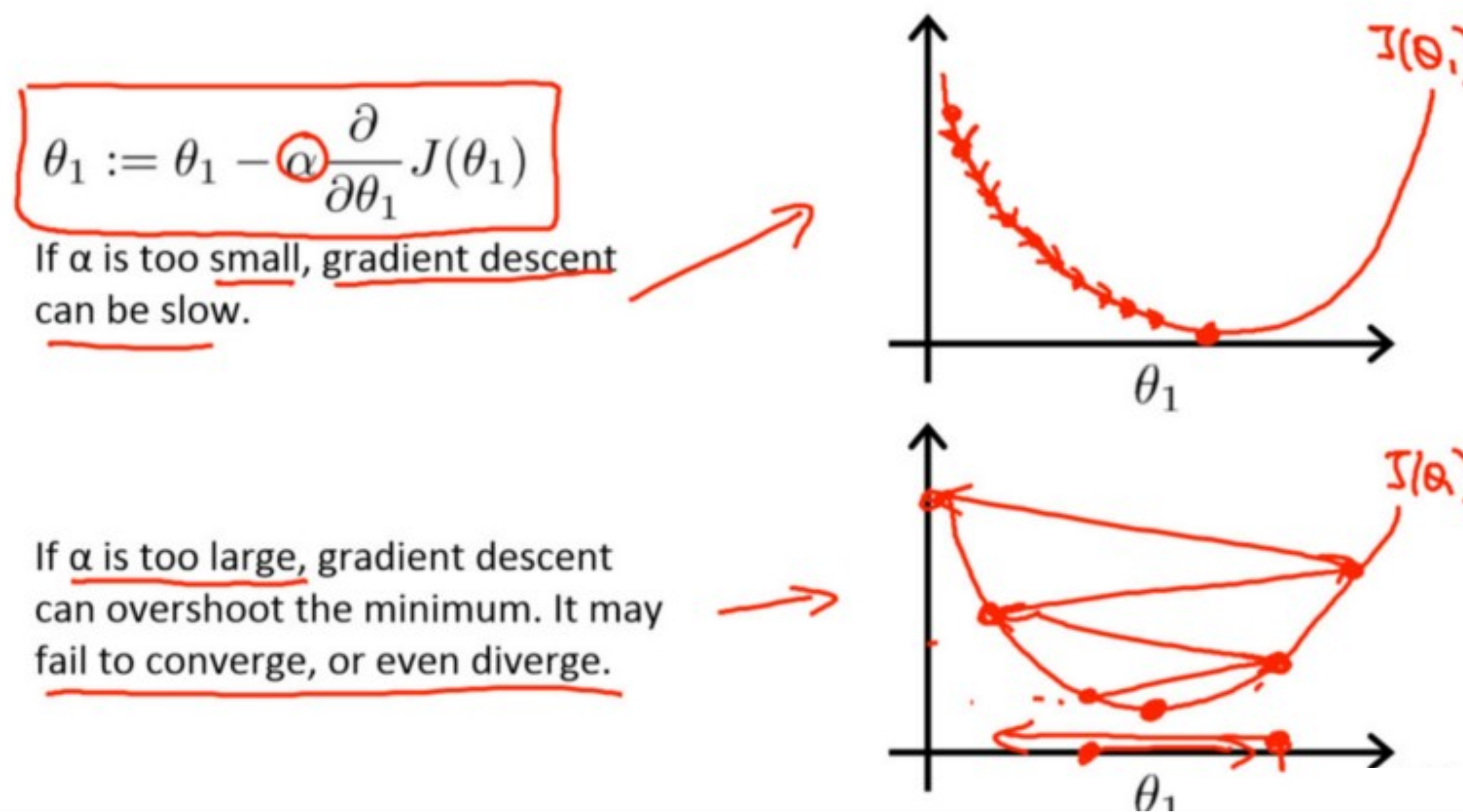
正規化 Regularization	隨機移除 Drop out
批次標準化 Batch Normalization	客製化損失函數 Customized Loss Function
回呼 Callback	提前終止 Early Stopping

本日知識點目標

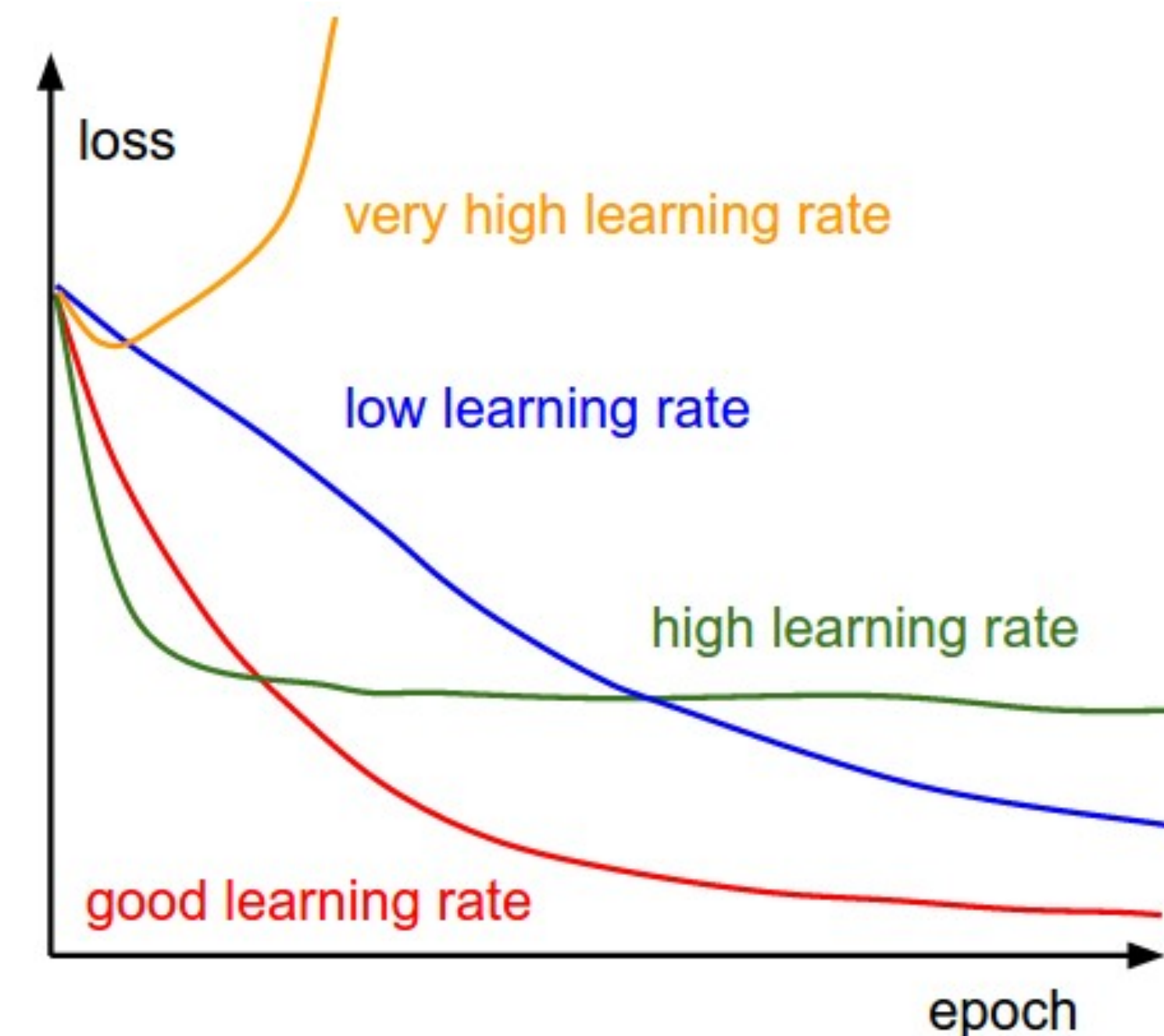
- 了解 Learning Rate 對訓練的影響
- 了解各優化器內，不同的參數對訓練的影響

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。



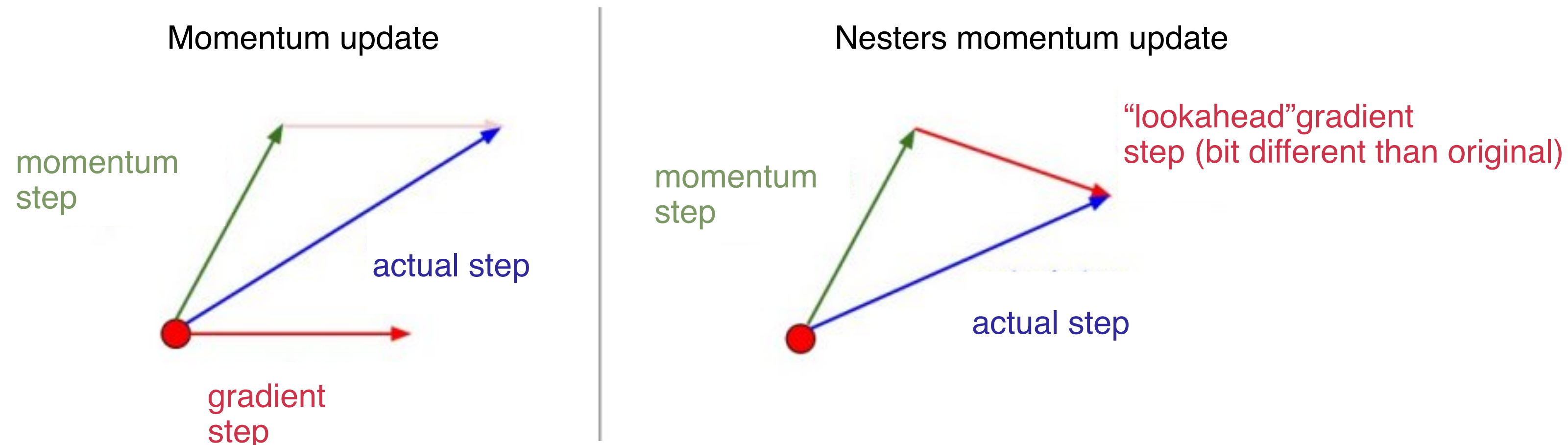
圖片來源：towardsdatascience.com



圖片來源：[cs231n.github](https://github.com/cs231n)

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。
 - $\text{Actual step} = \text{momentum step} + \text{gradient step}$
- Nesterov Momentum：拔草測風向
 - 將 momentum 納入 gradient 的計算
 - Gradient step computation is based on $x + \text{momentum}$



- 學習率對訓練造成的影響
 - 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

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

