**Sample Code & 作業內容**

請參考範例程式碼Day079\_LearningRateEffect，請完成以下作業：

請比較 SGD optimizer 不同的 momentum 及使用 nesterov 與否的表現

作業請提交Day079\_HW.ipynb

**[今日百日馬拉松作業 : 注意事項]**

1. 今日作業記憶體需求較大, 請同學執行時記得 shutdown 其他執行中的 .ipynb 檔

2. 提醒同學們今日執行會比較久(幾分鐘到數小時不等)

[檢視範例](https://ai100-2.cupoy.com/samplecodelist/D79)

**參考資料**

**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，讓收斂狀態趨於穩

參考連結：

1. [**Estimating an Optimal Learning Rate For a Deep Neural Network**](https://towardsdatascience.com/estimating-optimal-learning-rate-for-a-deep-neural-network-ce32f2556ce0)
2. [**cs231n: learning and evaluation**](http://cs231n.github.io/neural-networks-3/)
3. [**知乎-深度學習超參數簡單理解>>>>>>learning rate,weight decay 和 momentum**](https://zhuanlan.zhihu.com/p/23906526)