

ML Assignment XI

Unknown Question with Gemini

Name : 林欣妤
Student number : 314652023

November 25, 2025

1 Learning Rate 如何選取？

問題背景：在訓練神經網路時，我們利用梯度下降（Gradient Descent）來更新參數。學習率決定了我們每次往梯度相反方向（損失下降方向）移動的步長大小。

問題核心：選取學習率是一個在「收斂速度」與「訓練穩定性」之間的權衡（Trade-off）。沒有一個通用的最佳數值，通常需要透過以下策略來選取：

- (i) 基本原則 (The Goldilocks Principle) 過大 (Too Large)：損失函數 (Loss) 會劇烈震盪，甚至發散 (Diverge)，導致無法收斂。過小 (Too Small)：收斂速度極慢，容易卡在 ** 局部最小值 (Local Minima) ** 或鞍點 (Saddle Point)，浪費訓練時間。適中：損失隨時間穩定下降。
- (ii) 常見選取策略網格搜索 (Grid Search) / 試錯法：最傳統的方法。通常從 10 的幕次開始嘗試，例如 $0.1, 0.01, 0.001, 0.0001$ 。觀察最初幾個 Epoch 的 Loss 變化曲線，選擇下降最快且不震盪的數值。學習率衰減 (Learning Rate Decay / Scheduling)：開始時使用較大的學習率（如 0.1 ）以快速接近最佳解，隨著訓練進行（Epoch 增加），逐漸減小學習率（例如每 30 個 Epoch 降為原來的 $1/10$ ），以便在最佳解附近進行微調。使用自適應優化器 (Adaptive Optimizers)：現代深度學習常使用如 Adam 或 RMSProp 優化器。它們會根據梯度的歷史資訊自動為每個參數調整學習率。雖然仍需設定一個初始學習率（通常 Adam 設為 $1e^{-3}$ 或 $1e^{-4}$ ），但比傳統 SGD 更不敏感。
- (iii) 進階技巧：Learning Rate Finder 這是 Leslie Smith 提出的一種技巧（常見於 fast.ai 課程）。方法：在訓練初期，讓學習率從非常小（如 $1e^{-7}$ ）指數級增長到非常大（如 1 or 10 ）。觀察：繪製“Loss vs. Learning Rate”的圖表。選取：選擇 Loss 下降最陡峭（斜率最大）的那一段區間的中點作為初始學習率，而不是 Loss 最小的點（因為那裡已經快要發散了）。

參考資料：

- Deep Learning Book (Goodfellow et al.) - Chapter 8: 介紹了優化算法的基本原理。
- CS231n: Setting Hyperparameters: 史丹佛大學課程筆記，詳細解釋了如何透過圖表判斷學習率好壞 (Baby-sitting the learning process)。

2 Logistic Regression 與 Gaussian Discriminant Analysis (GDA)

都是用於分類，它們在模型類型上有什麼本質區別？

解答：

Logistic Regression 是判別式模型 (Discriminative Model)：它直接學習 $P(Y|X)$ ，即給定特徵 X 後屬於類別 Y 的機率。GDA 是生成式模型 (Generative Model)：它試圖模擬數據的分佈 $P(X|Y)$ 和 $P(Y)$ 。透過貝氏定理 (Bayes' Theorem)，它可以用來計算後驗機率 $P(Y|X)$ 。小知識：當數據量無限大且完全符合高斯分佈假設時，GDA 效率較高；但 Logistic Regression 對於錯誤的模型假設更具 Robustness。

參考資料：<https://cs229.stanford.edu/notes2020spring/cs229-notes2.pdf>