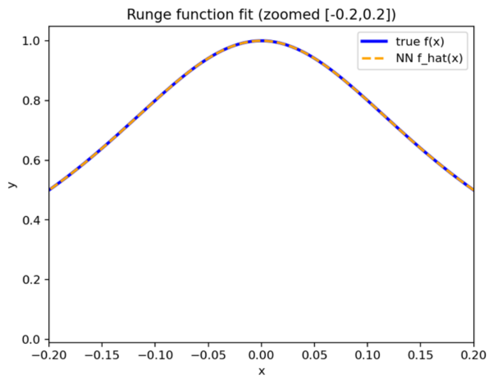
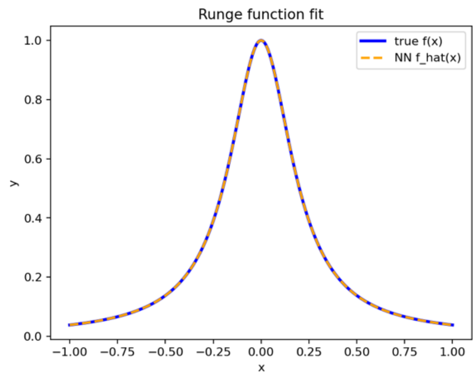
# 111652041 蔡翔宇 ML_HW3 程式報告

1. 在 HW2 中我們已經使用 NN:

   1(input layer)→64(hidden layer)→64(hidden layer)→1(output layer)，來逼近我們要的連續函數：$\frac{1}{1+25x^2}$，然後使用了 MSE(Loss function)來訓練我們的模型。(以下是我們超成功訓練)




而

2. 但 HW3 我們不只要逼近我們的連續函數 f(x) = $\frac{1}{1+25x^2}$ 還要想辦法來逼近他的導數 f'(x) = $\frac{-50x}{(1+25x^2)^2}$ ，並將其納入訓練與評估，讓模型不僅要學會逼近函數值，也要學會逼近函數的變化率。

3. 我們首先先建立 MLP 類別，而非固定，允許彈性設定層數、神經元數、激活函數，讓我們在不同結構下更加的靈活。就例如說 Activation function 我們可以使用不同的激活函數 (tanh,RELU,GELU,Sigmoid function)，並能測試不同設定下的效果。這種設計更符合研究實驗的需求，也能驗證 Universal Approximation Theorem 在不同架構下的效果。

4. 而在我們資料產生方面則以函數 make dataset 系統化產生 Train/Val/Test，並且測試集使用均勻網格取樣，確保曲線繪圖與導數檢驗更平滑、更有代表性。 不同於我們在 HW2 僅隨機生成樣本並切成 60/20/20。

5. 而我們是怎麼做出導數的呢？神經網路本身是一堆線性層 ＋ 非線性激活函數，也就是：

   $\hat{f}(x;\theta) = W_2\,\sigma(W_1 x + b_1) + b_2$, 這裡$\theta$代表所有權重。

   每個線性層和激活函數都是「可微分的」。所以我們可以利用auto grad（自動微分系統），讓電腦自動幫我們算出 $\hat{f}'(x) = \frac{\partial \hat{f}(x;\theta)}{\partial x}$ 。

6. 而我們現在的 Loss function 更加複雜了，我們不只要$MSE_{\mathrm{f}}$甚至還需要$MSE_{\mathrm{df}}$，因為我們還需要確保函數的微分也訓練的很相像。所以我們現在的 Loss function 定義為：

   Loss=$MSE_{\mathrm{f}} + \lambda_{\mathrm{i}} \cdot MSE_{\mathrm{df}}$ , where $\lambda_{\mathrm{i}}$ 是我們的權重。

比較「網路導數」和「真實導數」我們同時計算兩種誤差：

1. 函數值誤差：$\mathrm{MSE}_f = \frac{1}{N}\sum_i(\hat{f}(x_i) - f(x_i))^2$
2. 導數的誤差：$\mathrm{MSE}_{df} = \frac{1}{N}\sum_i(\hat{f}'(x_i) - f'(x_i))^2$

(1) 當 $\lambda_d = 0 \rightarrow$ 等於只看函數值誤差（跟 HW2 相同）。

(2) 當 $\lambda_d > 0 \rightarrow$ 同時考慮函數值與導數的擬合

在訓練資料裡，我們不只存 $f(x)$，還存 $f'(x)$，讓神經網路同時學「函數值」和「導數」！

7. 我們遇到的挑戰函數誤差 vs 導數誤差的量級不匹配，$MSE_{\mathrm{f}}$與 $MSE_{\mathrm{df}}$ 的數值尺度不同，直接相加會讓其中一項主宰學習。而我的做法是：
   (1) 對導數做 標準化/縮放（除了看 max $|f'(x)|$ 還有看訓練集的標準差）。
   (2) 用權重 $\lambda_d$ 做損失平衡；實務上先用小值（如 0.1），觀察兩項梯度範圍再調。
   (3) 兩階段訓練：先只訓練 $\mathrm{MSE}_f$，穩定後再開啟 $\lambda_d > 0$。

8. 邊界行為與「Runge 現象」帶來的斜率尖銳區學不穩，挑戰：端點附近變化率大，若採樣稀疏，導數學不好、曲線會劇烈抖動。而我的做法是：

   (1) 重點採樣：加密端點附近的點（或使用類 Chebyshev 的節點分佈）。
   (2) 測試用 均勻網格，訓練可用「均勻 + 端點加權」混合。
   (3) 在邊界加入 導數監督（端點附近多放帶 $f'(x)$ 的樣本）。

9. 導數資料可能不均衡（中間區域 vs 兩端區域）因為函數在中間平緩、兩端變化大，若取樣均勻，導數目標分佈不均。

   (1) 分段採樣配額：中央少取、兩端多取；或對導數大的樣本給更高權重！
   (2) 監看分區 MSE：把區間切段（端點/中段）各自計算誤差，對症調整採樣！