

國立雲林科技大學資訊管理系
Department of Information Management National Yunlin University
of Science & Technology

機器學習

專案作業一
Project assignment one

張詠晴 B11023063
林榮顯 B11023038
林界宏 B11023053
李浩誠 B11017010

指導教授：許中川 博士
Advisor：Chung-Chian Hsu, Ph.D.
中華民國 114 年 4 月
April 2024

摘要

本研究採用 Python 及 Keras/TensorFlow 實作前饋式神經網路，探討不同網路架構與超參數設定對模型在類別與數值預測方面的影響。研究主要涵蓋三項分析。首先，透過比較淺層與深層神經網路，評估隱藏層數量與模型深度對效能的影響。其次，針對隱藏層之啟動函數選擇，分析 Sigmoid、Softplus 與 ReLU 三種函數對模型表現的影響。最後，探討不同批次大小（batch size）與訓練週期（epochs）設定對模型收斂速度與準確度的影響。研究結果顯示，神經網路架構、啟動函數選擇及超參數設定皆對模型效能具有顯著影響，適當的參數調整可有效提升模型準確性與泛化能力。本研究之實驗結果與資料公布於 https://github.com/lin-roger/hw1_B11023038。

關鍵詞：深度神經網路、淺層神經網路、啟動函數、批次大小。

壹、緒論

一、動機

隨著人工智慧和機器學習技術的快速發展，神經網路已成為許多領域中廣泛應用的工具，特別是在圖像處理、自然語言處理和預測分析等領域。然而，隨著神經網路架構的複雜性提高，如何選擇適當的網路結構及超參數設置對於模型的表現至關重要。淺層神經網路（Shallow Neural Networks）和深度神經網路（Deep Neural Networks）在不同層數、節點數及激活函數等設定下，對預測性能的影響值得深入探討。因此，本研究旨在分析神經網路模型的架構設計及超參數選擇，並探索其對預測準確度和訓練時間的影響。

二、目的

本研究的主要目的是深入分析淺層神經網路與深度神經網路的訓練與測試性能，並探討不同設置對模型的影響，具體包括：

- **Q1**：分析淺層神經網路和深度神經網路的架構設計（隱藏層數量及每層節點數量），以及其對訓練資料集和測試資料集預測績效的影響。透過這一分析，期望了解不同網路深度和寬度如何影響模型的學習效果和準確度。
- **Q2**：分析不同激活函數（Sigmoid、Softplus、ReLU）在隱藏層中的應用，並探討它們對訓練資料集與測試資料集預測績效的影響。這將有助於了解激活函數選擇對神經網路模型性能的作用。
- **Q3**：探討不同的批次大小（batch size）與訓練輪數（epoch）設置，對訓練時間及模型預測績效的影響。這一部分的目的是研究如何平衡訓練效率與預測準確性，從而提供最佳的訓練設置。

通過這些分析，本研究期望為神經網路模型的設計與優化提供理論基礎，並對實際應用中如何選擇合適的網路結構與參數配置提供參考。

貳、方法

一、實作說明

Q1、Q2 與 Q3 都是先載入需要使用的套件與資料集，接著依據資料集類型執行對應的預處理，接著將需比較的參數傳入 `train_network()` 函數，令其訓練並紀錄不同參數之神經網路性能表現，最後再利用繪圖套件將其可視化。各檔案細節如下：

- **Q1**：載入資料集，做預處理(reshape 和 astype)然後設定隱藏層，使用不同層數與神經元數，對模型做訓練，最後於測試集計算指標，得出結果。
- **Q2**：載入資料集，做預處理(reshape 和 astype)然後設定隱藏層，在隱藏層中設定不同的啟動函數，對模型做訓練，最後於測試集計算指標，得到最終不同啟動函數的結果，在互相比較。
- **Q3**：載入資料集，做預處理(reshape 和 astype)然後設定隱藏層，在訓練中給予不同的 batch size 跟 epoch，得到結果再互相比較。

二、執行方法

下載 github 上之訓練權重以及程式，接著安裝所需套件（可使用 dev container 協議快速建構環境），再使用 Jupyter 開發環境讀取檔案的路徑，確認檔案可成功讀取後，即可執行程式。

參、實驗

一、資料集

(一)、Modified National Institute of Standards and Technology(MNIST)

MNIST 資料集由 70,000 筆灰階手寫數字相片組成，其中 60,000 筆為訓練集，剩下 10,000 筆為測試集。每張圖片由 28×28 個像素組成，每個像素以 0~255 的整數值表示灰度。

(二)、The Boston Housing

The Boston Housing 資料集由 506 筆紀錄美國麻薩諸塞洲波士頓地區的住房資訊組成，其中包含：人均犯罪率、平均房間數、一氧化氮濃度、房屋平均價格等共 13 個欄位組成(其中一個欄位因涉及種族歧視問題而被移除)。

二、前置處理

為加快模型收斂速度，本研究對 MNIST 資料集進行 Min-max normalization，對每個像素除以灰度最大值 255，確保灰度值在 $[0, 1]$ 的區間中。由於神經網路無法直接處理二階張量，本研究將 28×28 個像素組成的圖片展平為 784 維的向量。而為了讓模型處理離散的類別標籤，本研究使用 one-hot encoding(獨熱編碼)將其映射為 $|C|$ 維向量， $|C|$ 代表類別的數目，該向量是稀疏的，除了語類別對應的維度是 1 外，其他維度的值皆為 0。

而為確保模型不受不同欄位間數值分佈的差異影響，本研究對波士頓房價資料集進行 z-score normalization，確保每個欄位的均值為 0，標準差為 1。

三、實驗設計

本研究中所有模型皆使用 Adam(lr rate = $1e-3$, $\beta_1=0.9$, $\beta_2=0.99$, $\epsilon=1e-7$) 作為優化器，並固定從訓練集隨機選取 20% 的樣本作為驗證集。而所有在 MNIST 資料集進行訓練的模型皆使用類別交叉熵(Categorical Cross-entropy)作為損失函數，在 The Boston Housing 資料集訓練的模型則使用平均平方誤差(Mean Squared Error, MSE)作為損失函數。在性能指標的選擇，在 MNIST 資料集本研究使用準確率(Accuracy)，而 The Boston Housing 則使用平均絕對誤差(Mean Absolute Error, MAE)、平均絕對百分誤差(Mean Absolute Percentage Error, MAPE) 和均方根誤差(Root Mean Squared Error, RMSE)。而三種不同測試面向之實驗細節如下：

- **Q1**：分別測試了總神經元數(36, 72, 144, 288, 576, 1152, 2304)分別在一層、兩層和三層隱藏層的性能表現，每層皆使用 ReLU 作為啟動函數。皆是使用訓練期間驗證集損失最小的模型進行測試。
- **Q2**：固定使用三層隱藏層(512, 256, 128)之神經網路，並使用不同啟動函數(Sigmoid、Softplus、ReLU)於隱藏層，並觀察其性能變化。
- **Q3**：固定使用三層隱藏層(512, 256, 128)之神經網路，每層皆使用 ReLU 作為啟動函數。記錄不同 batch size 及 epochs 之訓練時間與性能表現，分別為 (batch size=128, epochs=10)、(batch size=8, epochs=10)、(batch size=128, epochs=50)。

表 1：不同神經元數與層數之神經網路性能比較表。

L	U/L	MNIST		The Boston Housing			
		Error ↓	Acc ↑	Error ↓	MAE ↓	MAPE↓	RMSE↓
1	2304	0.063	98.1	430.837	18.856	78.473	20.756
	1152	0.066	97.7	484.207	20.152	84.915	22.005
	576	0.064	98.1	526.610	21.127	89.845	22.948
	288	0.075	97.7	550.221	21.695	93.036	23.456
	144	0.094	97.3	574.329	22.173	95.182	23.964
	72	0.122	96.5	589.037	22.485	96.768	24.269
	36	0.164	95.3	609.238	22.917	99.101	24.679
2	1152	0.066	98.0	93.614	7.613	41.215	9.673
	576	0.069	97.9	100.999	8.348	42.684	10.046
	288	0.074	97.8	300.919	15.21	62.012	17.335
	144	0.082	97.4	455.841	19.297	79.563	21.346
	72	0.100	97.0	537.662	21.33	90.568	23.186
	36	0.138	95.9	577.450	22.258	95.787	24.028
	18	0.204	94.0	597.630	22.662	97.687	24.441
3	768	0.074	97.8	55.443	5.748	29.655	7.441
	384	0.076	97.8	100.228	7.825	45.750	9.991
	192	0.076	97.8	223.539	12.797	54.534	14.855
	96	0.086	97.5	464.253	19.530	80.938	21.542
	48	0.118	96.2	537.704	21.311	90.497	23.180
	24	0.159	95.2	588.473	22.462	96.563	24.255
	12	0.257	92.9	608.580	22.898	98.960	24.666

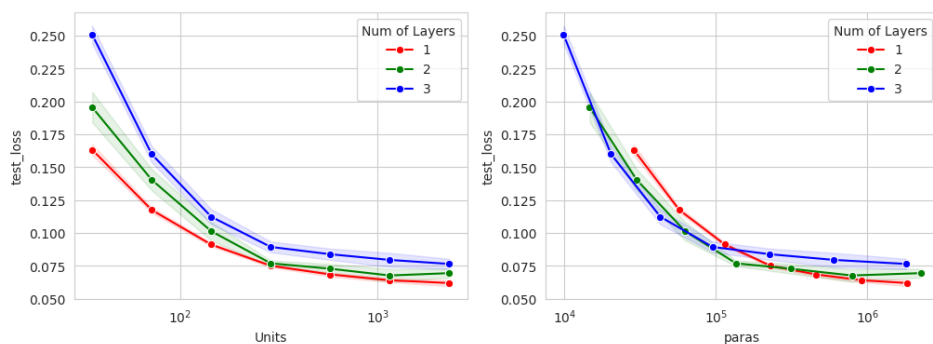


圖 1：不同層數與總神經元數（左）和總參數量（右）之神經網路於 MNIST 資料集之測試損失折線圖。

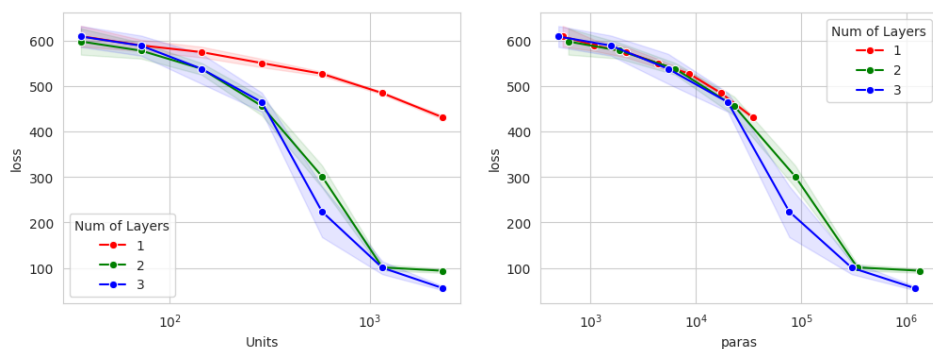


圖 2：不同層數與總神經元數（左）和總參數量（右）之神經網路於 The Boston Housing 資料集之測試損失折線圖。

四、實驗結果

不同神經元數量與深度之性能比較如表 1。在 MNIST 資料集中表現最好的是一層/每層有 2304 個神經元的淺網路，如圖 1 左。而在 The Boston Housing 資料集中表現最好的是三層/每層有 768 個神經元的深網路，如圖 2 左。觀察上述圖表可發現，在 MNIST 資料集中同等參數量下淺網路的性能略優於深網路，與現有理論不符，本研究推測深網路可能出現過擬和，使用正規化技術應該可以展現深網路的真實性能，不過礙於時間限制本研究沒有進行進一步的實驗。而在 The Boston Housing 資料集同等參數量下深網路的性能遠優於淺網路，與現有理論相符。

而不同啟動函數之性能比較如表 2，在 MNIST 資料集中表現最好的是 ReLU，在準確率領先 0.8。而在 The Boston Housing 資料集中表現最好的 Softplus，在 MAE 領先 2.295。觀察圖 3 與圖 4 可發現 Sigmoid 是全部啟動函數收斂速度最慢且性能最差的，因此本研究不建議使用 Sigmoid 作為隱藏層的啟動函數，而 ReLU 與 Softplus 分別在分類與回歸任務上表現了優異的性能與收斂速度，但可發現輕微的過擬和現象，搭配正規化技術可能可以進一步提升模型效能。

最後，不同訓練週期與批次大小之性能比較如表 3，在 MNIST 資料集中表現最好的參數組合是（週期=10, 批次大小=128），在準確率領先 0.1。而在 The Boston Housing 資料集中表現最好的(50, 128)，在 MAE 領先 0.006。觀察圖 5 與圖 6 可發現，參數組合為（週期=10, 批次大小=128）之模型在訓練後期已經陷入明顯的過擬和。而比較參數組合（週期=50, 批次大小=128）與參數組合（週期=10, 批次大小=8）之訓練時間可發現，週期與訓練時間成正比，批次大小與訓練時間成反比。

表 2：不同啟動函數之神經網路性能比較表。

Activation Function	MNIST		The Boston Housing			
	Error ↓	Acc ↑	Error ↓	MAE ↓	MAPE ↓	RMSE ↓
ReLU	0.076	97.8	99.487	7.750	46.740	9.971
Sigmoid	0.105	96.7	130.189	8.727	38.530	11.358
Softplus	0.100	97.0	54.641	5.455	31.130	7.383

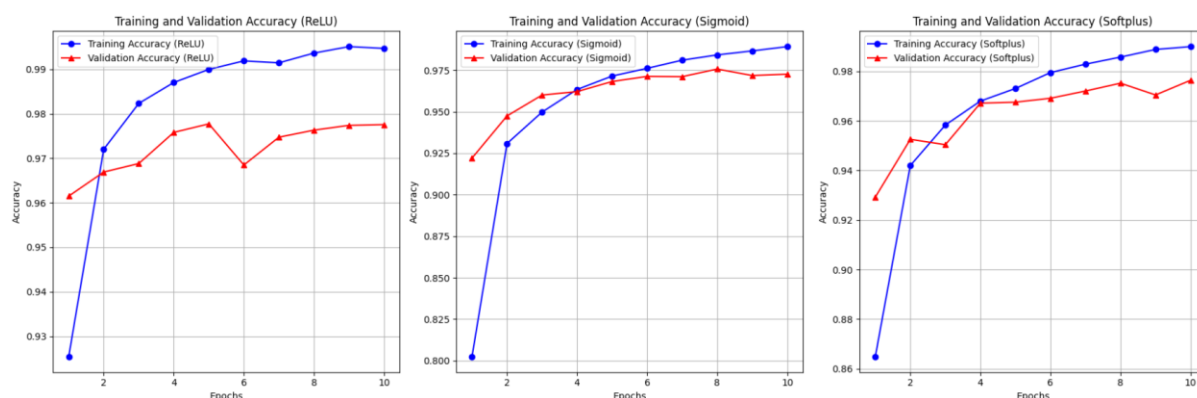


圖 3：不同啟動函數之神經網路於 MNIST 資料集之訓練/驗證準確率折線圖。

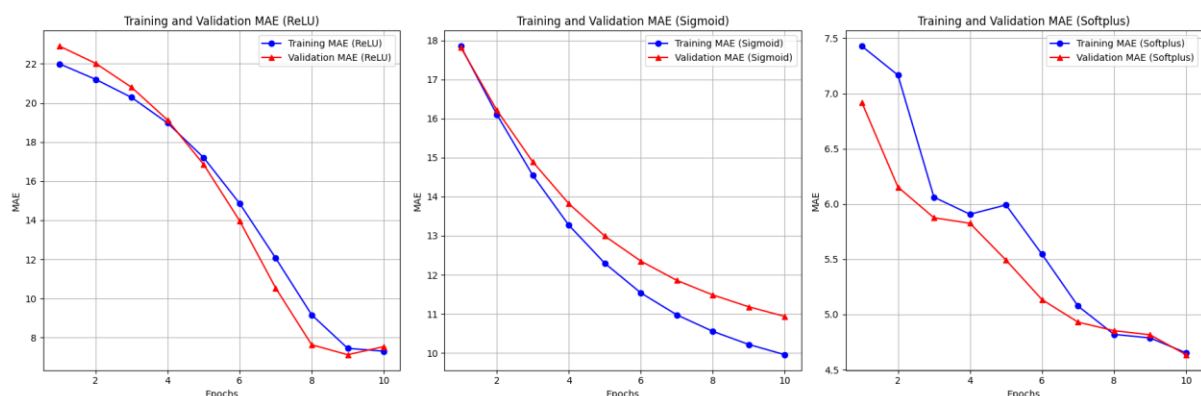


圖 4：不同啟動函數之神經網路於
The Boston Housing 資料集之訓練/驗證 MAE 折線圖。

表 3：不同訓練週期與批次大小之神經網路性能比較表。

Epochs	Batch size	MNIST			The Boston Housing				
		Error ↓	Acc ↑	Times↓ (s)	Error ↓	MAE ↓	MAPE↓	RMSE↓	Times↓ (s)
10	8	0.095	97.4	343.074	24.308	3.191	16.072	4.925	8.042
10	128	0.073	97.8	60.309	44.744	5.115	24.239	6.685	6.493
50	128	0.079	97.7	128.540	26.065	3.185	15.983	5.105	11.477

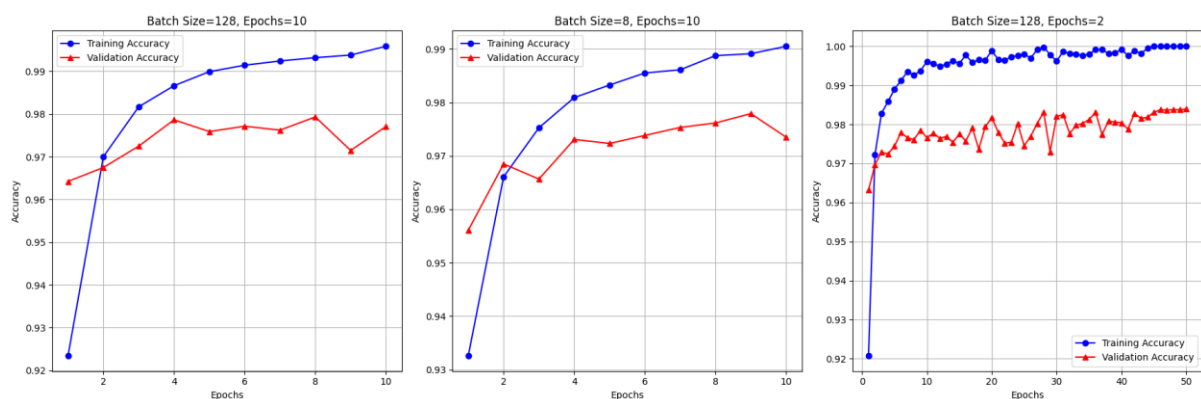


圖 5：不同訓練週期與批次大小之神經網路於
MNIST 資料集之訓練/驗證準確率折線圖。

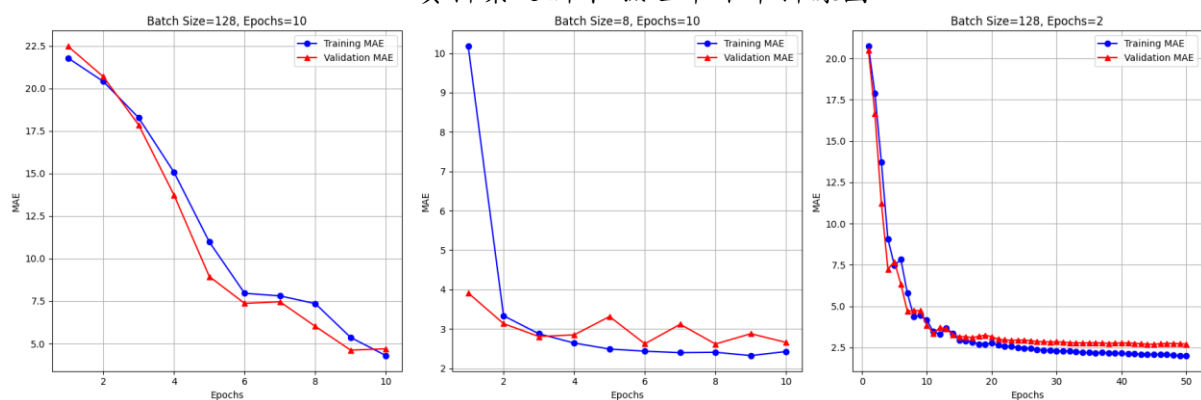


圖 6：不同訓練週期與批次大小之神經網路於
The Boston Housing 資料集之訓練/驗證 MAE 折線圖。

肆、結論

本研究對神經網路超參數組合：深度與神經元數、啟動函數、訓練週期數和批次大小共三種面向於分類任務與回歸任務進行測試，觀察並總結以下超參數選擇策略：

- **深度與神經元數**：深度且窄的神經網路優於淺且寬的神經網路，但在增加深度時配合使用正規化技術，否則匯出現過擬和的情況。
- **啟動函數**：ReLU 與 Softplus 分別在分類與回歸任務取得優異表現，並且擁有較快的收斂速度，Sigmoid 不具前述兩者的任何優點，不建議將其作為隱藏層的啟動函數。
- **訓練週期數和批次大小**：週期與訓練時間成正比，批次大小與訓練時間成反比。週期過大會導致過擬和。