



國立雲林科技大學

## 機器學習作業一

學號：[B11023222]

---

[前饋式神經網路預測類別與數值]

學 生：[周俊佑]

指導教授：許中川 教授

中華民國 114 年 3 月 17 日

# 摘要

本研究以 Python 及 Keras/TensorFlow 實作前饋式神經網路，用於類別與迴歸預測問題。實驗選用 MNIST 資料集進行手寫數字辨識，以及 Boston Housing Price 資料集進行房價預測。分析不同的激活函數以及 batch size 與 epoch 對預測結果的影響以及在不同隱藏層數量、神經元數量及激活函數設定下之預測誤差，亦參考了 MAE、RMSE 與 MAPE 三項指標。實驗結果表明，深層神經網路能提高正確率，但運算成本更高；而 ReLU 激活函數大致優於 Sigmoid 與 Softplus；調整 batch size 及 epoch 則可以對正確率與訓練速度進行抽象互換。隱藏層的數量與神經元的數量也確實影響了模型的預測績效。兩層隱藏層的模型通常具有較好的 RMSE，比單層的模型略優，但在某些情況下，多層模型可能反而降低了效能。這可能是因為過深的模型會導致過擬合，導致測試集的誤差增加。

## 一、緒論

### 1.1 動機

透過實際的前饋式神經網路預測類別與數值實驗來檢驗模型中各項參數，如 batch size、epoch、激活函數、隱藏層與神經元的數量等，對於分類任務的準確率與迴歸模型的預測表現所造成的影響及實際效果。為了同時涵蓋分類與迴歸兩種不同的機器學習問題。

本研究選擇了 MNIST 和 Boston Housing Price 這兩個標準資料集作為實驗對象。MNIST 資料集是一個經典的手寫數字辨識數據集，能夠測試神經網路在影像分類上的表現，並分析不同超參數對準確率的影響。另一方面，Boston Housing Price 資料集則是一個典型的房價預測數據集，適合評估神經網路在迴歸任務上的預測能力，並比較不同架構對 MAE、RMSE、MAPE 等誤差指標的影響。這兩個資料集不僅涵蓋不同類型的機器學習任務，也因為其適中的數據規模，使得訓練與測試過程能夠在合理的計算時間內完成，有助於更全面地分析神經網路的性能表現與最佳化方式。

### 1.2 目的

本研究的主要目的是透過前饋式神經網路在分類與迴歸問題上的應用，探討不同模型架構與超參數設定對預測績效的影響。為此，實驗分別選用 MNIST 手寫數字資料集進行影像分類，以及 Boston Housing Price 資料集進行房價預測，藉此分析神經網路在不同任務中的表現。

本研究將重點關注以下幾個面向：首先，探討隱藏層數量與神經元數量對模型訓練效果與預測準確度的影響，分析淺層與深層神經網路在不同數據集上的優勢與限制。其次，評估不同激活函數（Sigmoid、Softplus、ReLU）在模型學習過程中的影響，並比較其在分類與迴歸問題中的適用性。此外，研究亦將

探討 batch size 與 epoch 的不同組合，分析其對模型訓練時間與預測表現的影響，以尋找最適合的訓練策略。最後，透過 MAE（平均絕對誤差）、RMSE（均方根誤差）與 MAPE（平均絕對百分比誤差）等評估指標，客觀比較不同模型架構與超參數設定下的預測效能，進一步了解深度學習在實際應用中的優勢與挑戰。

## 二、方法

本研究採用 Python 與 Keras/TensorFlow 框架建立前饋式神經網路模型，分別針對分類任務（使用 MNIST 資料集）與迴歸任務（使用 Boston Housing Price 資料集）進行實驗，探討不同超參數設定對預測績效的影響。以下整理實驗方法內容：

### 2.1 資料前處理：

- MNIST 資料集：
  - 將 28x28 灰階圖像重構為 784 維一維向量。
  - 將像素值由 0~255 正規化至 0~1 區間。
  - 將標籤進行 one-hot 編碼（例如，將數字「3」轉換為 [0, 0, 0, 1, 0, 0, 0, 0, 0, 0]）。
- Boston Housing Price 資料集：
  - 計算每個特徵的均值與標準差，使用 Z-score 方法對數據進行標準化（轉換成均值為 0、標準差為 1 的標準正態分布）。
  - 資料中無缺失值，因此不需進行缺失值補值或刪除處理。

### 2.2 模型架構設計：

- MNIST 分類模型：
  - 輸入層：784 維向量（展平後的 28x28 圖像）。
  - 隱藏層：一層包含 1000 個神經元，激活函數使用 ReLU。
  - 輸出層：10 個神經元（對應 10 個分類），激活函數使用 softmax。
- Boston Housing Price 迴歸模型：
  - 輸入層：13 個神經元（對應 13 個特徵）。
  - 隱藏層：根據實驗設計，採用 1、2 或 3 層隱藏層，神經元數量可調（例如 32、64 或 128），各層激活函數可選用 ReLU、Sigmoid 或 Softplus。
  - 輸出層：1 個神經元，用於預測房價（不使用激活函數）。

### 2.3 模型訓練與評估：

- MNIST 分類任務：
  - 損失函數：Categorical Crossentropy。
  - 優化器：RMSprop。
  - 評估指標：準確率（Accuracy）。

- 實驗設計：比較兩組 epoch 值（5 與 10）與兩組 batch size（64 與 128）的組合，記錄測試準確率及訓練時間，並以表格與雙 Y 軸圖形呈現結果。
- Boston Housing Price 迴歸任務：
  - 損失函數：均方誤差（MSE）。
  - 優化器：RMSprop。
  - 評估指標：MAE（平均絕對誤差）、RMSE（均方根誤差）、MAPE（平均絕對百分比誤差）。
  - 實驗設計：透過改變隱藏層數量、神經元數量與激活函數的組合（固定 batch size 為 16 與 100 個 epoch），評估模型在迴歸任務中的表現，並以表格及折線圖呈現各組配置下的評估結果。

## 三、實驗

### 3.1 資料集

本研究使用兩個標準資料集：

- MNIST 資料集：包含 60,000 筆訓練資料與 10,000 筆測試資料，每筆資料為 28x28 像素的灰階圖像，共對應 10 個類別（0 至 9）。
- Boston Housing Price 資料集：包含 506 筆房價數據，每筆數據由 13 個特徵構成，描述影響房價的各項變數，目標是預測房價。

### 3.2 前置處理

對 MNIST 資料集，先將圖像資料 reshape 為一維向量，並將像素值正規化至 0~1；標籤部分則進行 one-hot 編碼。

對 Boston Housing Price 資料集，先計算各特徵的均值與標準差，並利用 Z-score 標準化方法將數據轉換至標準正態分布，確保各特徵尺度一致。由於資料中無缺失值，故不需額外處理。

### 3.3 實驗設計

在 MNIST 分類任務中，透過變動 epoch（5 與 10）及 batch size（64 與 128），探討其對模型測試準確率與訓練時間的影響。實驗流程包括：建立模型、記錄訓練時間、評估測試準確率，並將結果整理成表格與圖形呈現；圖形中以雙 Y 軸分別展示準確率與訓練時間。

在 Boston Housing Price 迴歸任務中，設計多組模型配置以比較隱藏層數量、神經元數量與激活函數（ReLU、Sigmoid、Softplus）的影響，並固定 batch size 與 epoch。模型訓練採用交叉驗證，最終利用 MAE、RMSE 與 MAPE 三項指標對模型預測誤差進行量化評估。各項實驗配置與結果將以表格整理，並以折線圖等方式直觀呈現各組合的差異。

### 3.4 實驗結果

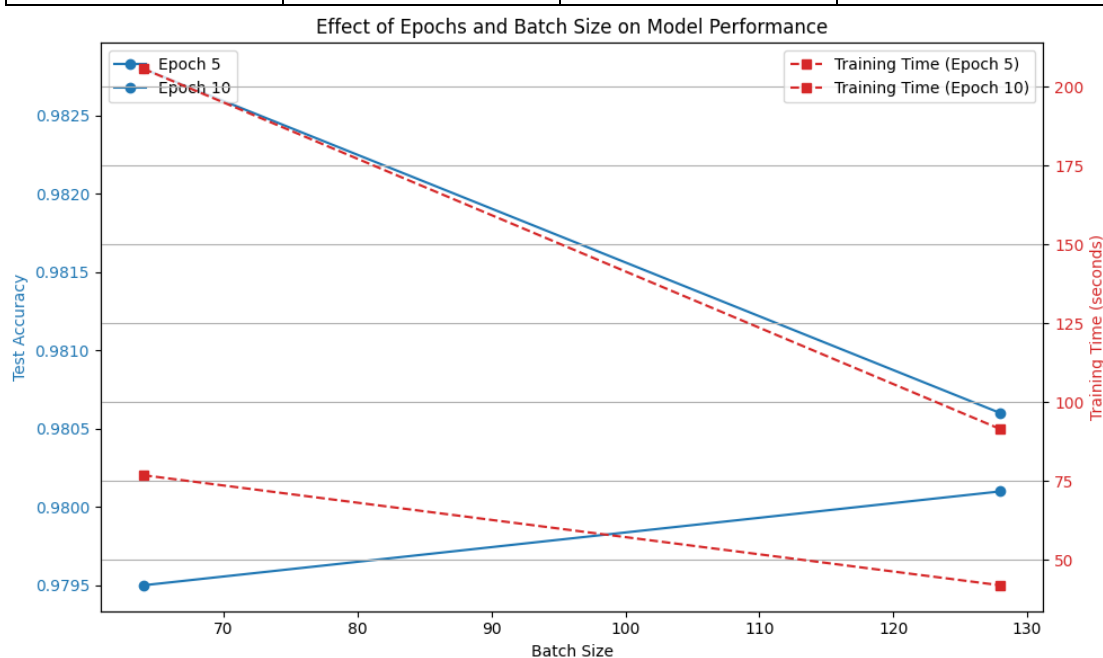
在 MNIST 分類實驗中，透過記錄不同 epoch 與 batch size 組合下的測試準確率與訓練時間，結果顯示：

- 當 epoch 為 5 時，batch size 為 64 的測試準確率約 0.9795，訓練時間約 76.84 秒；batch size 為 128 時，準確率約 0.9801，但訓練時間僅約 41.91 秒。
- 當 epoch 為 10 時，batch size 為 64 的測試準確率提升至 0.9828，訓練時間增加到約 205.80 秒；而 batch size 為 128 時，準確率約 0.9806，訓練時間約 91.46 秒。

從圖形與表格分析可看出，增大 epoch 會略微提升準確率，但同時顯著增加訓練時間；較大的 batch size 則能加快訓練速度，但可能略微影響準確率。

表一、實驗一 Mnist 手寫數字辨識實驗參數設定、成果表

Epoch	Batch Size	Accuracy	Training Time
5	64	0.9795	76.8s
5	128	0.9800	41.9s
10	64	0.9828	205.7s
10	128	0.9806	91.4s



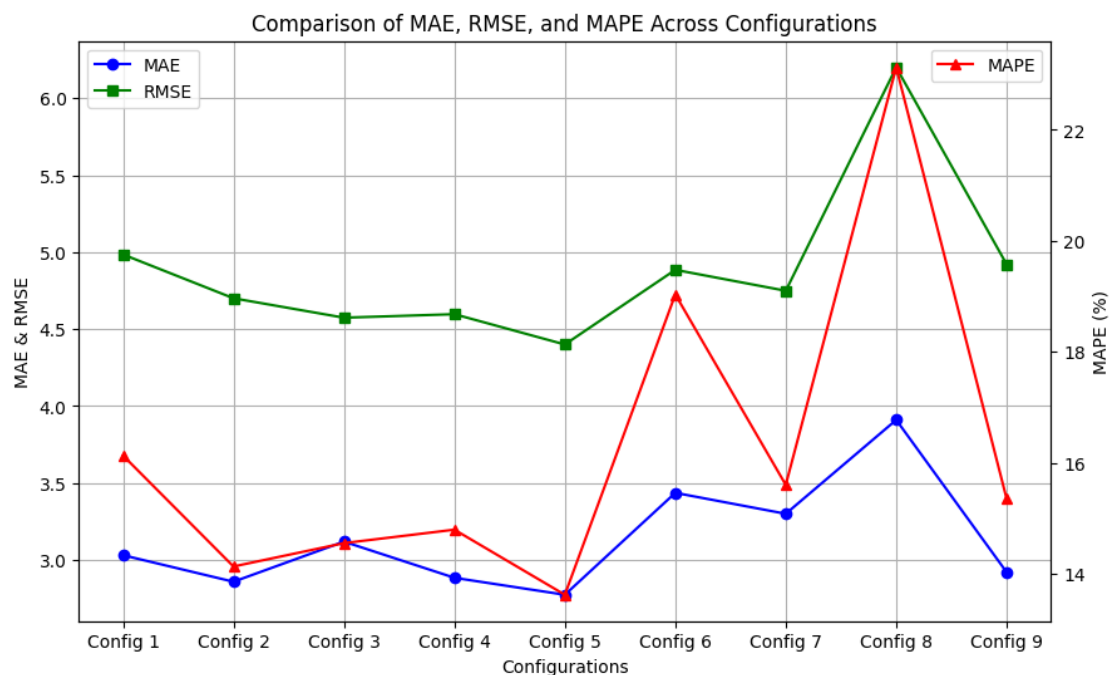
圖一、實驗一 Mnist 手寫數字辨識參數比較圖表

在 Boston Housing Price 迴歸實驗中，依據不同模型配置（變動隱藏層數量、神經元數量與激活函數），利用 MAE、RMSE 與 MAPE 指標進行評估。經過多組實驗比較發現，使用 ReLU 激活函數的模型通常能取得較低的預測誤差，而在隱藏層數量與神經元數量上，兩層隱藏層的配置（例如 2 層、64 個神經元）往往能取得較優的 RMSE 值，但過深的模型（例如 3 層）可能因過擬合而降低測試表現。各組配置的具體指標值整理如下表：

表二、實驗二 Boston Housing Price 迴歸實驗參數設定、成果表

設置	MAE	RMSE	MAPE
----	-----	------	------

{'hidden_layers': 2, 'units': 32, 'activation': 'relu', 'batch_size': 16, 'epochs': 100}	3.031	4.985	16.13%
{'hidden_layers': 2, 'units': 64, 'activation': 'relu', 'batch_size': 16, 'epochs': 100}	2.859	4.701	14.14%
{'hidden_layers': 2, 'units': 128, 'activation': 'relu', 'batch_size': 16, 'epochs': 100}	3.119	4.574	14.56%
{'hidden_layers': 1, 'units': 64, 'activation': 'relu', 'batch_size': 16, 'epochs': 100}	2.884	4.597	14.80%
{'hidden_layers': 3, 'units': 64, 'activation': 'relu', 'batch_size': 16, 'epochs': 100}	3.436	4.885	19.04%
{'hidden_layers': 2, 'units': 64, 'activation': 'sigmoid', 'batch_size': 16, 'epochs': 100}	3.300	4.749	15.61%
{'hidden_layers': 2, 'units': 64, 'activation': 'softplus', 'batch_size': 16, 'epochs': 100}	3.910	6.200	23.12%



圖二、實驗二 Boston Housing Price 迴歸實驗參數影響指標比較圖

## 四、結論

本研究透過前饋式神經網路在 MNIST 與 Boston Housing Price 兩個標準資料集上的實驗，探討了不同超參數（包括隱藏層數量、神經元數量、激活函數、batch size 與 epoch）對模型預測績效與訓練時間的影響。實驗結果顯示，對於分類任務，隨著 epoch 的增加，模型準確率有小幅提升，但訓練時間也明顯延長；較大的 batch size 可大幅縮短訓練時間，但可能略微影響預測

效能。對於迴歸任務，使用 ReLU 激活函數的模型通常能取得較低的 MAE、RMSE 與 MAPE，而適當調整隱藏層數量與神經元數量（例如兩層 64 個神經元）有助於降低預測誤差，過深的模型可能會因過擬合而使測試誤差上升。

綜上所述，本研究驗證了超參數設定對前饋式神經網路在分類與迴歸任務中的重要影響，並指出在不同應用場景下應根據需求在預測效能與訓練效率間取得平衡。未來可進一步探討其他模型架構或更多超參數組合，以提升模型在實際應用中的表現。