

國立雲林科技大學

資訊管理研究所

機器學習

專案作業一

陳杰龍

M11323002

國立雲林科技大學資訊管理研究所

湯曜謙

M11323010

國立雲林科技大學資訊管理研究所

鄒彥平

M11323049

國立雲林科技大學資訊管理研究所

中華民國一一四年三月十九日

## 摘要

本專案旨在透過實際操作，深入探討不同神經網路架構、超參數設置以及資料集特性對模型訓練與預測效能的影響。研究的核心目標包括探討隱藏層數量、神經元數量、激活函數、批次大小（batch size）和訓練週期（epoch）對分類（如 MNIST）與回歸（如 Boston Housing Price）問題的影響。實驗結果顯示，不同的隱藏層數量和神經元配置對模型的效能有顯著影響，並且較小的批次大小雖然能夠在初期達到較高準確度，但較大的批次大小能提高訓練效率和穩定性，並改善模型的最終預測結果。激活函數方面，ReLU 在分類任務中表現最佳，而 Softplus 則在回歸任務中表現最優。

關鍵字：神經網路、隱藏層、激活函數、批次大小、訓練週期、MNIST、Boston Housing Price、分類、回歸

## 壹、緒論

### 一、動機

隨著深度學習技術的迅速發展，神經網路已在影像識別、語音處理及數值預測等領域取得顯著突破。對於機器學習領域的學習者而言，通過實際操作並應用這些技術，能夠有效掌握其核心原理。本專案旨在利用 Python 及 Keras/TensorFlow 架構前饋神經網路模型，深入探討不同神經網路架構、超參數設定以及資料集特性對模型訓練與預測效能的影響。透過這些實驗，將不僅提升模型設計、訓練與評估的能力，更為未來在處理更為複雜的問題時，奠定堅實的深度學習應用基礎。

### 二、目的

本研究的目的是透過實際操作和分析，深入探討神經網路架構、激活函數及訓練參數(如 batchsize 和 epoch)在不同機器學習問題(包括分類與回歸問題)中的預測效能影響，期望通過這些分析，能夠優化神經網路模型的結構與參數，從而提升其在分類和回歸問題中的預測準確度。具體目標包括：

1. 探討淺層與深層神經網路在不同資料集(如 MNIST 和 BostonHousingPrice)上的表現差異。
2. 分析隱藏層激活函數(如 Sigmoid、Softplus、ReLU)對訓練與測試資料集預測準確度的影響。
3. 研究不同 batchsize 和 epoch 設定對訓練時間及模型預測效能的影響。

## 貳、方法

### 一、程式架構

第一個程式使用人工合成的分類資料集來訓練神經網路，並比較不同激活函數的效果，架構包括資料集生成與標準化、多層神經網路模型的建立(含兩層隱藏層與 Dropout 防止過擬合)、訓練與測試，並將結果繪製成圖表比較。第二個程式基於 MNIST 數據集進行手寫數字分類，探索不同批次大小和訓練週期對訓練時間、準確度及損失的影響，並繪製出相關圖表。第三個程式針對波士頓房價數據集進行回歸分析，計算 MAE、RMSE 和 MAPE 等誤差指標，並探討不同批次大小與訓練週期對模型表現的影響，最後以圖表形式展示結果。

### 二、程式的方法

將程式檔案保存為 .py 或 .ipynb 檔案，並確保安裝了所需的依賴庫，如 TensorFlow、NumPy 和 Matplotlib。對於 .py 檔案，可以在命令行中使用 python

<filename>.py 來執行；而對於.ipynb 檔案，則可以在 Jupyter Notebook 環境中直接運行。在執行程式之前，請確保相關數據集（如 MNIST、波士頓房價等）已正確下載並可用，並且所有必要的套件已經安裝。

## 參、緒論

### 一、資料集

在本研究中，選用了兩個廣泛應用的標準資料集來進行分類與回歸問題的實驗。這些資料集分別為 MNIST 手寫數字資料集與 BostonHousing 房價預測資料集。MNIST 資料集通常用於圖像分類任務，而 BostonHousing 資料集則用於回歸預測，特別是對房價的預測。這些資料集在機器學習領域中具有重要的代表性，並廣泛用於算法性能的測試與比較。

MNIST 資料集是用於手寫數字識別的經典資料集，包含 60,000 張手寫數字圖像作為訓練資料，和 10,000 張圖像作為測試資料。每張圖像的解析度為 28x28 像素，並以灰度值(範圍從 0 到 255)表示。每個圖像對應一個標籤，該標籤為 0 到 9 之間的整數，表示該圖像所對應的數字。由於每張圖像包含 784 個像素，因此在進行神經網路訓練前，需要將其重塑為一維向量。此資料集的主要用途為數字分類，並可用於訓練前饋式神經網路進行多類別分類問題。該資料集具有以下特點：

- 訓練集：60,000 張圖像
- 測試集：10,000 張圖像
- 每張圖像包含 784 個像素(28x28=784)
- 標籤為 0 到 9 的類別標籤

表 1 MNIST 手寫數字資料集(部分資料)

編號	像素值(第一行)	像素值(第二行)	...	標籤
1	0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0	...	0
2	0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0	...	1
3	0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0	...	2
...	...	...	...	...

BostonHousing 資料集旨在預測波士頓市各地區的房價，並包含 13 個特徵變數，這些變數涵蓋了如犯罪率、住宅區面積、學校質量等多個社會經濟指標。資料集的目標變數為房價中位數。該資料集共有 404 筆訓練資料與 102 筆測試資料，用於訓練及評估回歸模型。此資料集的主要用途為回歸預測，並可用於訓練回歸模型來預測房價。該資料集具有以下特點：

- 訓練集：404 筆資料
- 測試集：102 筆資料
- 13 個特徵變數，包括犯罪率、土地面積、學校質量等
- 目標變數為該地區的房價中位數，數值型資料

表 2 BostonHousing 房價預測資料集(部分資料)

編號	CR IM	ZN	IN DU S	CH AS	NO X	R M	AG E	DI S	RA D	TA X	PT RA TI O	BL OS S	M ED V
1	0.0 063 2	18. 0	2.3 1	0	0.5 38	6.5 75	65. 2	4.0 9	1	296 .0	15. 3	396 .9	24. 0
2	0.0 273 1	0.0	7.8 7	0	0.4 69	6.4 21	78. 9	4.9 7	2	242 .0	17. 8	396. 9	21. 6
3	0.0 272 9	0.0	7.8 7	0	0.46 9	7.18 5	61. 1	4.9 7	2	242 .0	17. 8	394. 0	34. 7
...	...	...	...	...	...	...	...	...	...	...	...	...	...

## 二、前置處理

資料前置處理是模型訓練過程中的關鍵步驟，旨在將原始資料轉換為適合神經網路模型處理的格式。在本研究中，對於 MNIST 資料集，我們進行了圖像重塑並將像素值進行了標準化處理；對於 BostonHousing 資料集，則進行了均值與標準差的標準化處理，確保資料具備一致性並提升模型訓練的效果。這些前置處理步驟對提高模型的預測準確度和訓練效率起到了至關重要的作用。

在處理 MNIST 資料集時，首先將每張 28x28 像素的二維圖像轉換為 784 維的一維向量，這是因為神經網路通常處理的是一維向量，而非二維陣列。經過此轉換後，訓練集的圖像被重塑為 (60000,28\*28) 的形狀，測試集則重塑為 (10000,28\*28) 的形狀。為了提升模型的學習效果，我們對圖像進行了數據標準化處理，將像素值從原本的 0 到 255 歸一化至 0 到 1 之間。這樣的處理不僅能加速神經網路的訓練過程，還能防止某些特徵的數值範圍過大，從而主導模型的學習。最後，標籤(即數字 0 到 9)採用了類別標籤編碼(one-hotencoding)轉換為向量形式，每個標籤被轉換為 10 維向量，其中每一維對應於數字 0 到 9 的位置，這樣有助

於進行多類別分類任務。

對於 BostonHousing 資料集，由於各特徵變數的數值範圍差異較大，因此我們對訓練集和測試集進行了標準化處理(Z-score normalization)。具體而言，首先計算訓練資料中每個特徵的均值和標準差，並將訓練資料中的每個特徵進行標準化，使其符合標準正態分佈(均值為 0，標準差為 1)。測試資料則使用訓練資料中計算得到的均值和標準差進行標準化處理，以避免資料外洩的問題。這樣的處理方法確保了不同特徵之間的數據對模型的影響是均等的，有助於提升回歸模型的預測準確性。由於 BostonHousing 資料集中的特徵大多為連續型數據，因此進行標準化處理是十分必要的，能夠有效地幫助神經網路理解並學習各特徵之間的內在關聯。

### 三、實驗設計

#### 1. 隱藏層數量及節點數量之影響—MNIST 手寫數字資料集

在這個實驗中，我們的目的是探討模型中隱藏層的數量(層數)如何影響其測試準確度。根據這段程式碼，實驗測試的範圍是從 0 層到 11 層。每次改變層數時，模型的架構會隨之調整，層數越多，模型的複雜度和潛在學習能力也隨之增強。具體而言，當 num\_layers 設為 0 時，模型只有一個輸入層和一個輸出層，這是一個非常簡單的線性模型。而當 num\_layers 增加時，模型將會加入更多的隱藏層，這些隱藏層的神經元數量為 128 個，並且每一層使用 ReLU 激活函數進行非線性轉換，最後加入 10 個神經元的輸出層，使用 softmax 激活函數來進行分類。模型使用 rmsprop 優化器來進行訓練，損失函數選擇了 categorical\_crossentropy，適用於多分類問題。每次訓練後，我們都會計算訓練集和測試集的損失與準確度，並將結果收集起來。測試結果會繪製成圖，顯示隨著層數增加，測試準確度是否有所提升，並分析層數對模型效能的影響。

第二個實驗中，我們將專注於每個隱藏層中的神經元數量(層大小)對測試準確度的影響。具體而言，我們選擇了一個固定的隱藏層數量—3 層隱藏層，並且分別測試了每層隱藏層神經元數量為 8、16、32、64、128、256、512、1024 和 2048 等不同的數值。這樣的設定能夠幫助我們了解增加隱藏層中神經元數量是否會提高模型的表現，或是是否存在過度擬合的風險。每次測試，我們會將同樣的數量層數(3 層隱藏層)與不同的層大小結合，進行訓練。訓練過程與前一個實驗相同，使用 rmsprop 優化器，並且訓練 5 個時期。這樣的實驗設計可以讓我們比較不同層大小下，測試準確度的變化，從而了解隱藏層大小的選擇如何影響模型的泛化能力。

第三個實驗綜合了層數與每層神經元數量的組合來深入分析這些參數對測試準確度的影響。在這次實驗中，我們測試了層數從 1 層到 5 層不等，並且每層的神經元數量分別設為 10、32、64、128、256 等幾種數量的組合。這樣的組合

設置能夠幫助我們探索不同層數與不同隱藏層大小對模型表現的綜合影響，進一步了解如何調整層數和神經元數量來達到最佳的測試準確度。在這個實驗中，模型會根據每一組參數進行訓練，並且將每個配置的測試準確度、訓練損失和參數量進行比較分析。這樣的實驗設計可以揭示哪些參數組合最能提高測試準確度，並且讓我們對模型的結構選擇有更深入的理解。

## 2. 隱藏層數量及節點數量之影響—BostonHousing 房價預測資料集

實驗測試了隱藏層數量(從 1 層到 11 層)對模型準確度的影響。build\_model 函數中，num\_layers 參數控制了隱藏層的數量，每一層有 64 個神經元，並且使用 ReLU 激活函數。模型的損失函數為均方誤差(MSE)，並計算了 MAE、RMSE、MAPE 等多個評估指標。實驗中的過程是逐一增加層數，並計算測試數據集上的各項指標。隨後，這些指標會被繪製為圖形，並呈現不同層數對測試準確度的影響。

實驗研究了每層隱藏層神經元數量(layer\_size)對模型表現的影響。在這裡，模型固定為兩層隱藏層，但每層隱藏層的神經元數量會分別測試 32、64、128、256、512 等不同的數值。通過這樣的設置，我們可以檢查增加或減少隱藏層神經元數量對於模型訓練的影響。這個實驗的步驟包括訓練模型，計算不同配置下的 MAE、RMSE、MAPE 等評估指標，並且比較每個隱藏層大小的效果，並繪製出結果圖表。

在這個實驗中，我們綜合了層數和每層隱藏層神經元數量兩個參數。具體來說，我們測試了 1 層至 3 層的層數，每層神經元數量為 10、32、64、128 和 256 等不同的數值。這樣的組合可以幫助我們深入了解層數和每層神經元數量的選擇如何綜合影響模型表現。實驗的流程是針對每一組層數與神經元數量的組合，訓練並評估模型，並且將結果與其他組合的測試準確度進行比較。結果顯示在表格和圖形中，這樣可以幫助我們識別最適合該數據集的配置。

## 3. 激活函數之影響—MNIST 手寫數字資料集

這個程式的實驗設計是針對比較不同激活函數在二元分類任務中的效果。首先，程式生成了一個較為困難的分類資料集，並加入雜訊以提高任務的難度。資料集經過標準化後，分為訓練集與測試集。接著，程式建立了一個深度神經網絡模型，包含三層隱藏層，每層使用不同的激活函數 (sigmoid、softplus、relu) 並加上 Dropout 層來防止過擬合。訓練過程中，模型進行 100 週期的訓練，並計算訓練和測試準確度。實驗的目的是比較不同激活函數對訓練與測試準確度的影響，並通過繪製訓練過程中的驗證準確度曲線來分析這些激活函數的效果。

## 4. 激活函數之影響—BostonHousing 房價預測資料集

這個程式的實驗設計是針對回歸問題，並比較不同激活函數（Sigmoid、Softplus 和 ReLU）對模型性能的影響。首先，程式生成了一個帶有隨機噪音的回歸資料集，並對資料進行標準化處理。資料被劃分為訓練集和測試集，並使用三種不同的激活函數來訓練回歸模型。每個模型使用兩層隱藏層，每層包含 64 個神經元，並且使用均方誤差（MSE）作為損失函數，最終模型的預測結果會根據 MAE（平均絕對誤差）、RMSE（均方根誤差）和 MAPE（平均絕對百分比誤差）來評估。實驗結果會繪製三個圖表，分別顯示三種激活函數對 MAE、RMSE 和 MAPE 的影響。目的是比較不同激活函數在回歸任務中的表現，並分析其在誤差指標上的差異。

#### 5. Batchsize 與 epoch 之影響—MNIST 手寫數字資料集

程式的實驗設計是針對手寫數字識別任務，使用 MNIST 數據集來訓練一個簡單的神經網絡模型，並分析不同批次大小（batch size）和訓練週期（epochs）對模型表現的影響。程式中定義了三種不同的批次大小（16、32、64）和三種不同的訓練週期（30、60、90），每組設定會分別訓練模型並記錄訓練時間、測試準確度和測試損失。最終，程式會根據不同的批次大小和訓練週期繪製三個圖表，分別展示訓練時間、測試準確率和測試損失隨訓練週期變化的趨勢。這樣的實驗設計旨在探索批次大小與訓練週期如何影響模型的效能和訓練效率。

#### 6. Batchsize 與 epoch 之影響—BostonHousing 房價預測資料集

程式的實驗設計是針對波士頓房價預測任務，使用波士頓房價數據集來訓練一個深度神經網絡，並分析不同批次大小（batch size）和訓練週期（epochs）對模型表現的影響。程式中設定了三種批次大小（16、32、64）和三種不同的訓練週期（60、110、160），每組配置都會進行訓練，並記錄訓練時間、預測誤差（包括 MAE、RMSE、MAPE）。模型的設計包含兩層隱藏層，每層有 64 個神經元，激活函數使用 ReLU，並且輸出層預測房價。最終，程式會繪製四個圖表，分別展示訓練時間、MAE、RMSE 和 MAPE 隨訓練週期變化的趨勢，以評估不同參數配置對模型性能的影響。

### 四、實驗結果

#### 1. 隱藏層數量及節點數量之影響—MNIST 手寫數字資料集

如圖 1，顯示了不同層數的神經網絡在訓練和測試過程中的表現。從實驗結果可以看出，測試準確度在增加層數後達到穩定，並在 2 層時達到約 97% 的高準確度，隨後即使層數繼續增加，測試準確度也沒有顯著提升。訓練準確度隨層數增加而逐步提高，並維持在高水平，顯示訓練過程中模型適應性良好。儘管訓練



損失逐漸降低，測試損失在層數增多後出現波動，未顯示明顯改善。這表明，在這個實驗中，增加隱藏層數並未顯著提升測試準確度，並可能導致過擬合的風險。因此，較少的層數(如 2 層)即可達到理想的測試準確度，顯示出過多層數對模型效能的增益有限。

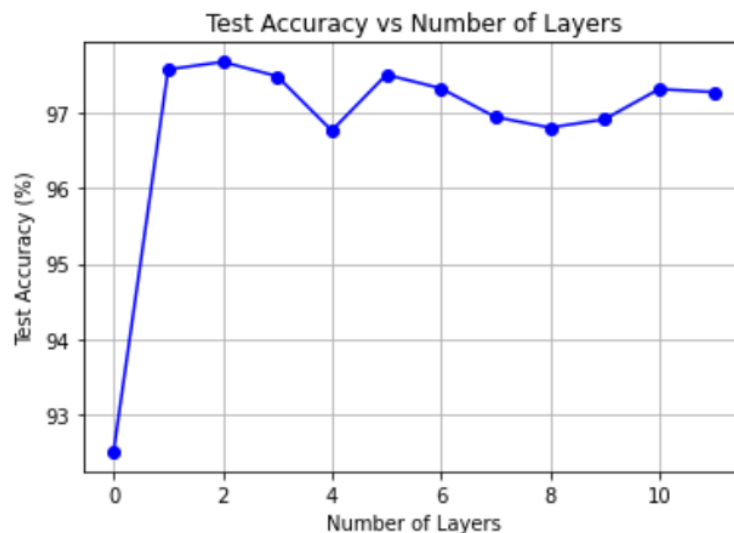


圖 1 Test Accuracy vs Number of Layers

如圖 2，顯示了不同隱藏層神經元數量對訓練和測試準確度的影響。隨著神經元數量的增加，測試準確度顯著提升，尤其是在神經元數量達到 128 個以上時，測試準確度穩定在約 97% 以上。當神經元數量超過 512 時，測試準確度進一步穩定，並沒有顯示出顯著的增長。訓練準確度隨著神經元數量的增加而逐漸提升，並在 512 個神經元以上達到穩定的高水平。整體來看，增加神經元數量有助於提高模型的測試準確度，但當神經元數量達到一定規模後，準確度的提升幅度變得有限，顯示出模型在達到一定規模後表現逐漸穩定。

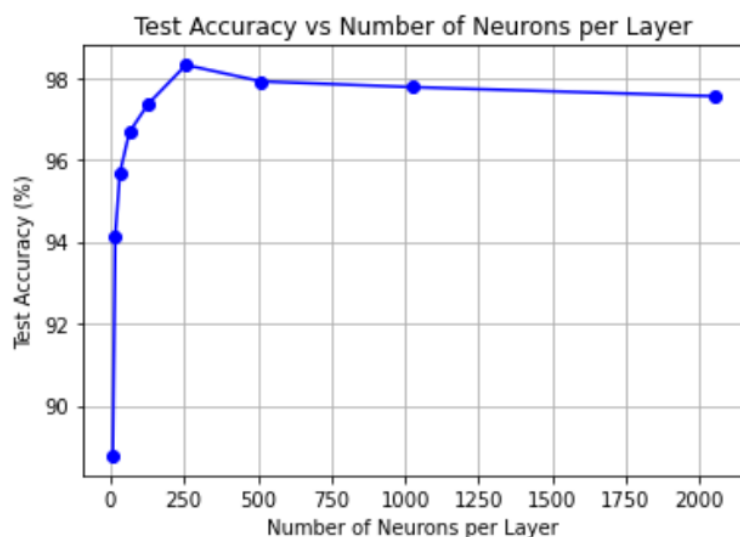


圖 2 Test Accuracy vs Number of Neurons per Layer

如圖 3，顯示了不同隱藏層數與每層神經元數量對測試準確度的影響。隨著

參數數量(即神經元數量的增加)，測試準確度逐漸上升，並且在某些情況下會達到穩定。特別是當隱藏層數為 1 層時，隨著參數數量的增多，測試準確度迅速提升，並在約 0.2 百萬個參數時達到最高，接近 98%。隨著隱藏層數的增加(2 層、3 層、4 層和 5 層)，測試準確度的增長趨勢逐漸變緩，尤其是當參數數量超過 0.2 百萬時，各層數的測試準確度已趨於穩定，並在 97% 以上，顯示在一定的參數數量下，增加隱藏層數對測試準確度的提升效果逐漸減少。這表明，增加神經元數量能顯著提升模型的表現，但當達到一定規模後，效果開始趨於平緩。

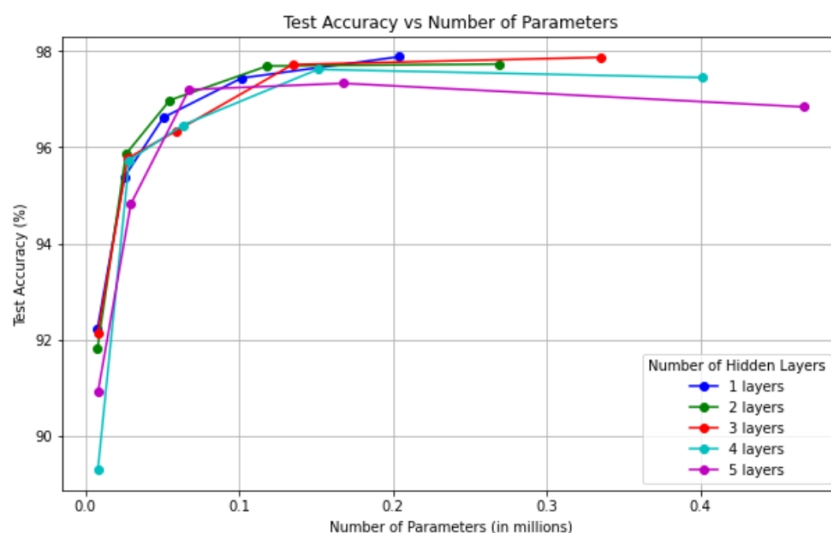


圖 3 Test Accuracy vs Number of Parameters

## 2. 隱藏層數量及節點數量之影響－BostonHousing 房價預測資料集

如圖 4，測試 MAE(平均絕對誤差)和 RMSE(均方根誤差)在不同層數的變化相對較小，但當層數增加到 5-6 層時，RMSE 有顯著下降的趨勢，表示模型的預測誤差有所減少。然而，MAPE(平均絕對百分比誤差)則在 50%至 53%之間波動，顯示模型的相對誤差沒有穩定下降的趨勢。整體而言，較低的 RMSE 和 MAE 出現在 5-6 層附近，這可能是模型的最佳結構範圍，進一步增加層數對於誤差的改善效果有限，甚至可能導致誤差波動增加。因此，在這個實驗中，適當選擇 5-6 層的模型可能是較好的選擇，以平衡預測準確性與模型複雜度。

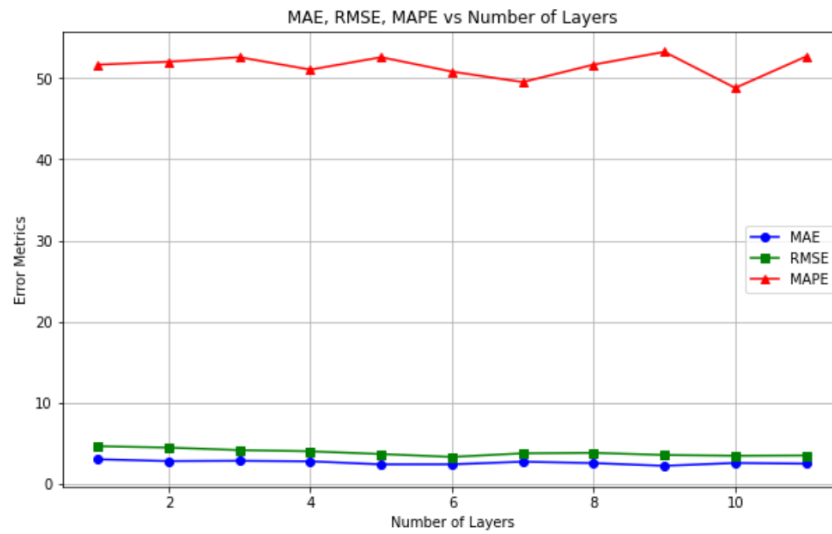


圖 4 MAE, RMSE, MAPE vs Number of Layers

如圖 5，隨著每層神經元數量的增加，測試 MAE(平均絕對誤差)和 RMSE(均方根誤差)呈現下降趨勢，這表示較大的神經元數量有助於提升模型的預測能力。其中，在 256 個神經元時，MAE 和 RMSE 均達到較低的值，而 MAPE(平均絕對百分比誤差)則在 512 個神經元時下降至最低點 49.37%。這顯示增加神經元數量可以有效降低誤差，但過多的神經元可能帶來計算成本增加的問題。綜合考量，256-512 個神經元的配置可能是較佳選擇，能在誤差與效能間取得平衡。

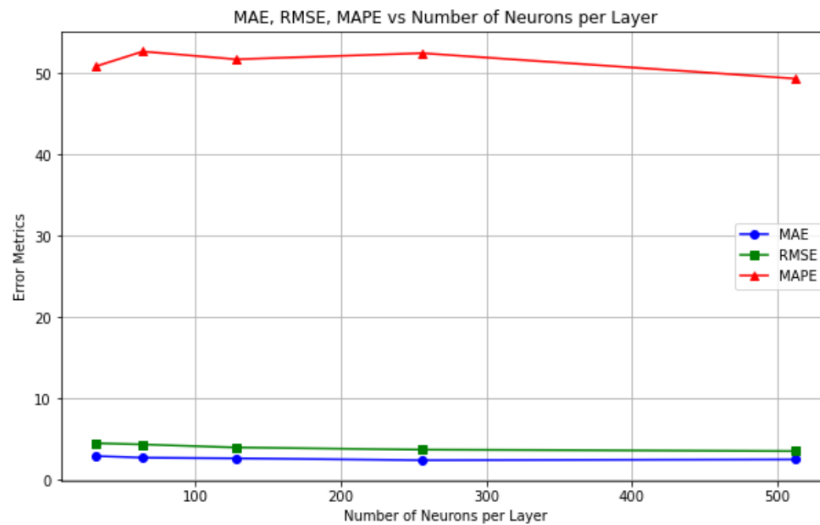


圖 5 MAE, RMSE, MAPE vs Number of per Layer

如圖 6，測試 MAE(平均絕對誤差)與 RMSE(均方根誤差)整體上隨著參數數量的增加而下降，這表示增加模型的複雜度有助於提升預測準確性。然而，不同層數的模型表現有所不同。例如，具有 3 層且 128 或 256 個神經元的模型在 MAE 和 RMSE 上達到了最低值，顯示較深層的網絡能有效學習數據模式。然而，MAPE(平均絕對百分比誤差)在不同的參數設置下呈現較大波動，特別是在 3 層 256 個神經元時 MAPE 反而上升至 54.19%，顯示較深層的模型可能在某些情況下導致過擬合或在比例誤差方面效果不佳。因此，根據這些結果，3 層 128 個神經元的模型可能是較好的選擇，因為它在 MAE 和 RMSE 上表現最佳，且沒有明顯的 MAPE 增加問題。

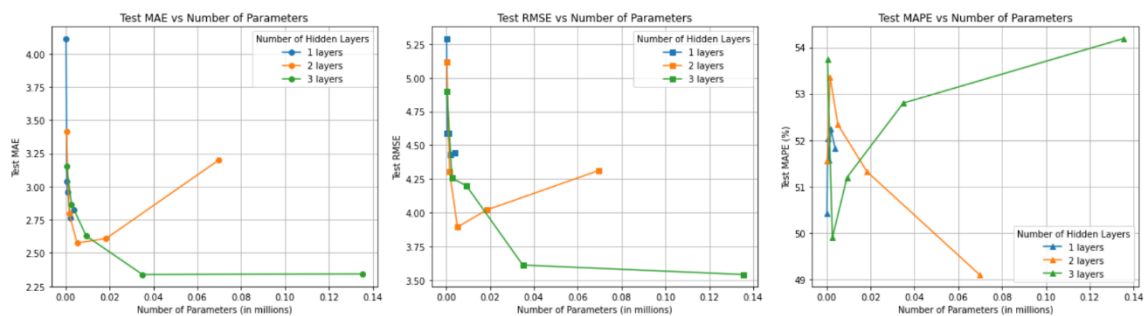


圖 6 MAE, RMSE, MAPE vs Number of Parameters

### 3. 激活函數之影響—MNIST 手寫數字資料集

如圖 7，不同的激活函數對驗證準確度的影響。隨著訓練週期的增加，使用 ReLU 激活函數的模型在大約 20 週期後表現出最佳的穩定性，並在後續的訓練中持續提高準確度，最終達到最高的驗證準確度。Softplus 函數的表現稍遜，雖然它在初期有較快的增長，但其最終的準確度略低於 ReLU。相比之下，Sigmoid 函數在整個訓練過程中的進展最為緩慢，並且最終達到的驗證準確度較低，顯示出在這個分類任務中，Sigmoid 函數的效果不如其他兩種激活函數。因此，從實驗結果來看，ReLU 是最適合此任務的激活函數，能夠在較短的時間內達到較高的驗證準確度，而 Sigmoid 則顯得較為滯後。

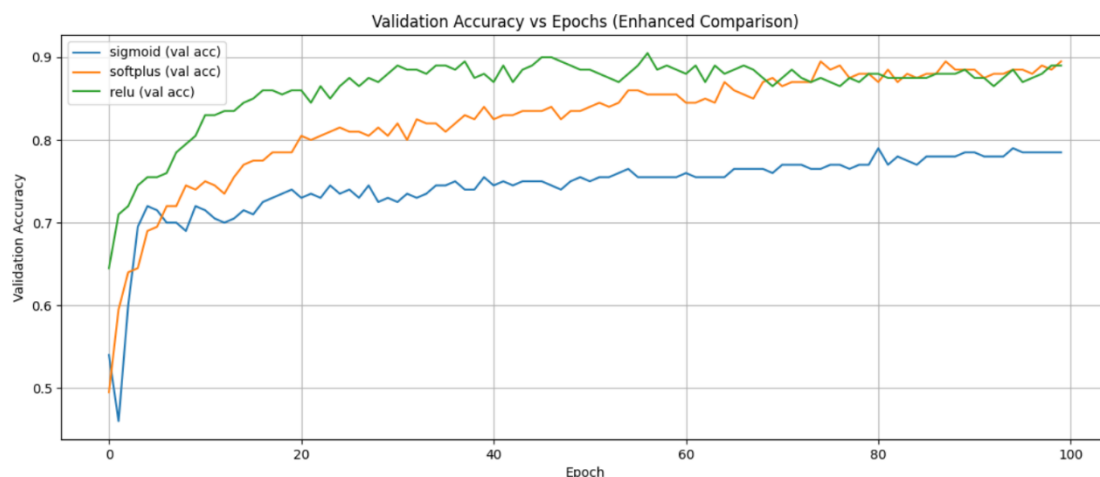


圖 7 Validation Accuracy vs Epochs (Enhanced Comparison)

#### 4. 激活函數之影響—BostonHousing 房價預測資料集

如圖 8，不同激活函數對回歸模型的三個誤差指標(MAE、RMSE 和 MAPE)有顯著影響。Sigmoid 激活函數的模型在所有三個指標中表現最差，無論是 MAE、RMSE 還是 MAPE，都達到了較高的數值。相比之下，Softplus 函數的模型在所有指標上表現最好，尤其是在 MAE 和 RMSE 上，它顯示出顯著的誤差降低。ReLU 函數的表現介於 Sigmoid 和 Softplus 之間，儘管其在某些指標上的誤差較低，但仍不如 Softplus。因此，從實驗結果來看，Softplus 激活函數在這個回歸任務中提供了最佳的性能，能夠有效降低預測誤差。



圖 8 Comparison of Activation Functions in Regression Model Performance

#### 5. Batchsize 與 epoch 之影響—MNIST 手寫數字資料集

如圖 9，批次大小對訓練時間、測試準確度和測試損失有顯著的影響。隨著訓練週期的增加，較小的批次大小（16）會導致較長的訓練時間，而較大的批次大小（64）則會使訓練時間相對較短。這是因為較小的批次大小需要更多的迭代來完成每個訓練週期，從而增加總訓練時間。關於測試準確度，較小的批次大小（16）在訓練初期表現較好，但隨著訓練週期的增加，準確度呈現下降的趨勢，而較大的批次大小（64）則能保持穩定的測試準確度。最後，測試損失隨著訓練週期的增加呈現上升趨勢，並且較大的批次大小（64）的損失值較低，表明其在訓練過程中可能更有效地收斂，儘管其準確度較低。因此，較小的批次大小雖然能在初期獲得較高準確度，但在更長的訓練期間，較大的批次大小可能更有利於模型的穩定性與收斂效果。

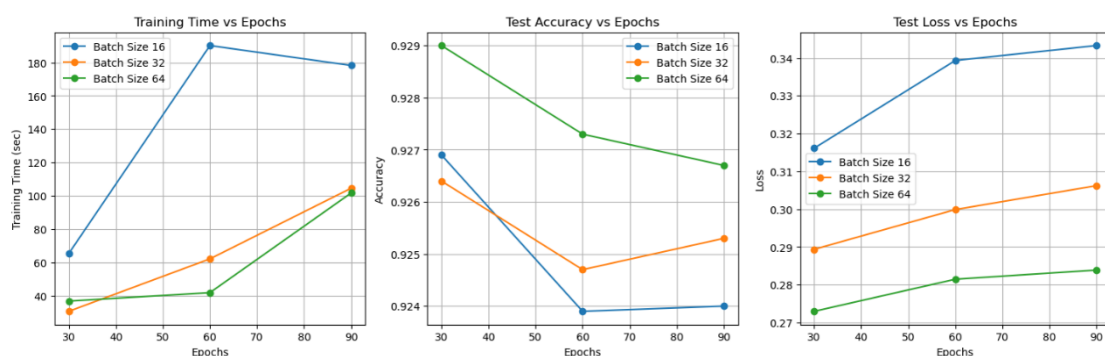


圖 9 Effect of Batch Size on Training Time, Accuracy, and Loss

## 6. Batchsize 與 epoch 之影響－BostonHousing 房價預測資料集

如圖 10，隨著批次大小（batch size）和訓練週期（epochs）的變化，模型的訓練時間、預測誤差（包括 MAE、RMSE 和 MAPE）均呈現不同的趨勢。首先，訓練時間隨著訓練週期的增加而增加，且較小的批次大小（16）在相同週期下訓練時間最長，較大的批次大小（64）訓練時間則相對較短。關於預測誤差，隨著訓練週期的增長，測試集上的 MAE、RMSE 和 MAPE 逐漸減少，顯示隨著更多訓練週期的進行，模型的表現有所改善。較小的批次大小（16）在所有指標上都表現得較差，而較大的批次大小（64）則能在較短的訓練週期內達到較低的預測誤差，顯示其在訓練效率和準確性上的優勢。這表明，較大的批次大小能提高模型的訓練效率，並且能夠在較短的時間內達到更好的預測性能。

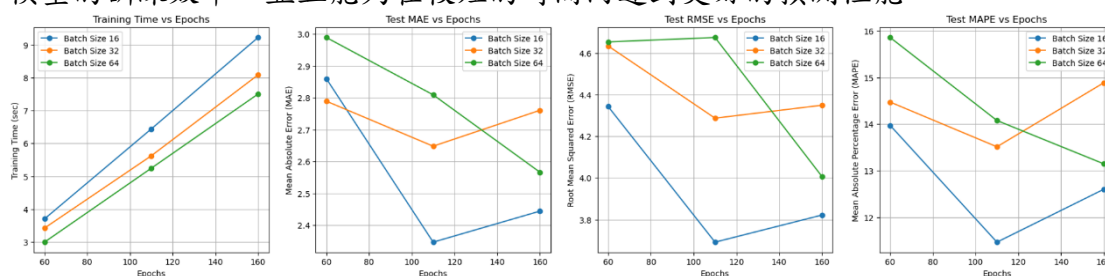


圖 10 Impact of Batch Size on Training Time Error Metrics (MAE, RMSE, MAPE)

## 肆、結論

本研究探討了不同神經網路架構與訓練設置對模型效能的影響。實驗結果表明，在分類任務中，隨著隱藏層數量的增加，測試準確度並未顯著提高，而在回歸任務中，增加層數和神經元數量有助於提升預測精度。較小的批次大小對訓練準確度有正面影響，但會導致較長的訓練時間，較大的批次大小則能有效提高訓練效率並降低誤差。激活函數的選擇對模型效能影響深遠，其中 ReLU 在分類任務中表現優越，而 Softplus 在回歸任務中最佳。因此，根據不同的任務特性，選擇適當的神經網路架構和訓練參數對於提升模型效能至關重要。