

Homework 2

Name: 謝伊婷

Student ID:112034533

1.

兩個 hyper-parameters : hidden units 和 batch size 。

Hidden units 設定為 128、256 和 512，batch size 設定為 16、32 和 64。

進行了九次實驗，結果以表格形式呈現，可以觀察這些 hyper-parameters 對訓練、驗證和測試數據集上的損失 (Loss) 和準確率 (Accuracy) 的影響。

Hidden Units	Batch Size	Train Loss	Train Accuracy(%)	Val Loss	Val Accuracy(%)	Test Loss	Test Accuracy(%)
128	16	0.36	84.127	0.482	76.543	0.139	77.419
128	32	0.363	84.656	0.444	81.481	0.07	77.419
128	64	0.5	74.074	0.512	80.247	0.048	61.29
256	16	0.344	88.889	0.54	77.778	0.133	67.742
256	32	0.366	84.127	0.485	81.481	0.071	67.742
256	64	0.48	75.132	0.561	69.136	0.046	61.29
512	16	0.31	87.831	0.512	77.778	0.12	77.419
512	32	0.36	84.656	0.541	70.37	0.07	70.968
512	64	0.447	77.249	0.58	71.605	0.043	67.742

(#選擇最後一個 Epoch 的值，計算 loss 和 accuracy)

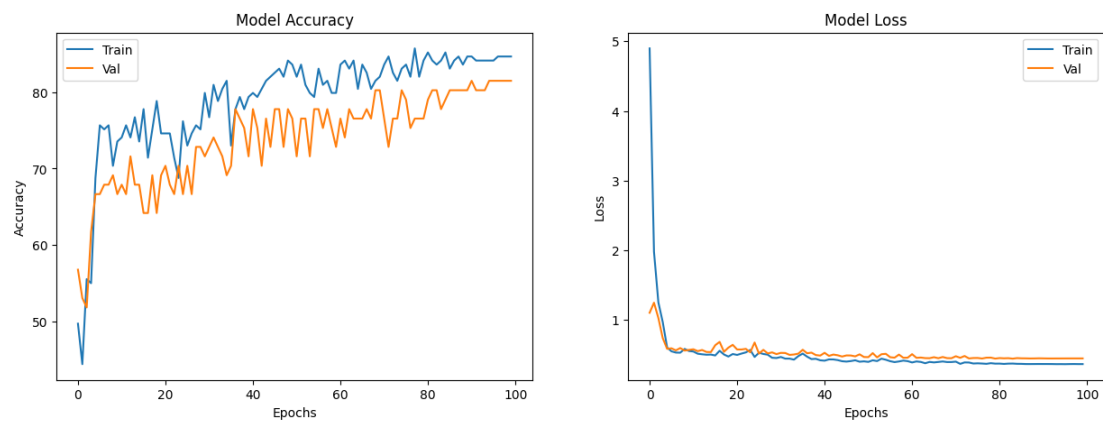
2.

Hidden units 為 128 和 batch size 為 32 的組合，在驗證集上有較高的 accuracy，在測試集上也有較佳的能力。

Hidden units 增加至 256 時，batch size 為 16 的設定雖然訓練集的 accuracy 最高，但在驗證和測試集上的提升有限，可能 overfitting。

Hidden units 增加至 512 時，當 batch size 為 64 時，驗證損失最大，可能有 overfitting。

綜合來看，hidden units 為 128 和 batch size 為 32 的設定，在三個 datasets 展現了穩定且較高的 performance，是本次實驗中表現最佳的 hyper-parameters 組合。附圖為此組參數設定下，在訓練和驗證階段，accuracy 和 loss 的視覺化。



3.

A discrepancy in accuracy between the training and test datasets，通常是模型在訓練數據集過擬合造成的。過擬合是模型學習到訓練數據中的 noise 和細節，而不是潛在的數據分佈。

此外，如果 train set 和 test set 的 distribution 不一致，或者 test set 包含了 model 之前未見過的新特徵和模式，也可能造成 accuracy 的差異。

為了減少這種差距，可以用正則化，例如 Dropout 或 weight decay，或是用 data Augmentation 和交叉驗證來增加 model 的泛化能力。

另外，數據預處理和特徵工程，使訓練集更好地代表 real-world 的 data 分佈，也是關鍵。

4.

Features 的選擇在 machine learning models 的建立中，扮演著關鍵角色，目標在於選出對預測目標有貢獻的重要 features，同時移除無關或冗餘的 features。

這可以透過不同的方法實現，包括統計測試（如卡方檢定）、模型基礎方法（例如利用樹模型來評估特徵重要性）和迴歸係數（如 Lasso 迴歸）。

特徵選擇能夠降低模型的複雜度，增強模型的泛化能力，減少過擬合的風險，並且可能降低計算成本。適當的特徵選擇不僅提高模型的準確率，還能使模型訓練與推斷過程更加高效。

在進行特徵選擇時，重要的是要通過交叉驗證來評估特徵子集的性能，以確保所選特徵的有效性。

(reference : ChatGPT)

5.

對於處理表格數據，除了人工神經網絡（ANNs）之外，一個較為適合的深度學習模型是深度學習決策樹（Deep Learning Decision Trees，DLDT）。

DLDT 結合了決策樹與深度學習的優點，特別為結構化數據設計。這種模型保留決策樹直觀、易於解釋的特性，同時引入深度學習能夠自動提取高層次特徵的能力。DLDT 通過多個層次，逐層深化特徵的判斷樹，使其能夠捕捉特徵間複雜的非線性關係。

DLDT 的關鍵特點包括其階層式特徵分解能力，這使得它在每個分支點考慮特徵的同時，能夠學習到特徵之間的深層次交互作用。此外，DLDT 模型還結合了嵌入式學習，可以進一步提取特徵表示，並將其用於更精細的分類決策。其優勢在於模型不僅能給出預測結果，還能提供決策路徑的解釋，這對於許多需要可解釋性的商業應用來說非常有價值。

(reference : ChatGPT)