

National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications Homework 2

Name: 黃品翰

Student ID: 112034511

1. Hyper-parameters : batch-size(32,64,128),learning rate(0.0001,0.0005,0.001)

	Batch Size	Learning Rate	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
0	32	0.0001	0.492796	75.661376	0.545519	71.604938	0.679079	61.290323
1	32	0.0005	0.376539	85.714286	0.501380	77.777778	0.656845	61.290323
2	32	0.0010	0.342501	86.772487	0.501939	77.777778	0.633524	77.419355
3	64	0.0001	0.542164	71.957672	0.552392	70.370370	0.688356	58.064516
4	64	0.0005	0.458156	81.481481	0.510056	74.074074	0.634137	58.064516
5	64	0.0010	0.408548	85.185185	0.460713	76.543210	0.654880	64.516129
6	128	0.0001	0.534587	75.132275	0.586614	70.370370	0.664219	64.516129
7	128	0.0005	0.474956	78.835979	0.537495	71.604938	0.638776	64.516129
8	128	0.0010	0.481516	75.661376	0.570937	71.604938	0.875166	58.064516

2. 在不同 batch size 的實驗中，整體來看 batch=32 時常能獲得較佳的訓練與測試準確率，可能由於較小的 batch size 能提供更頻繁的梯度更新並提升泛化能力。而在調整 learning rate 時，batch=32 與 64 下，較高的 learning rate 反而帶來更佳表現，但在 batch=128 時，learning rate 過大易造成訓練不穩定，最終表現不如預期。可能代表 batch size 與 learning rate 之間存在交互影響，須同時考量才能獲得最佳結果。從繪製的訓練與驗證曲線亦能看出此差異。

3. 這次實驗中，所有組合的訓練準確率皆高於測試準確率，顯示可能發生了 overfitting。由於訓練 epoch 數較多且未設置 Early Stopping，模型比較會記住 training data 的 pattern，導致在測試資料上的表現較差。其他可能原因像是資料量不足或分佈不一致、缺乏 L2 regularization、dropout，以及學習率和 batch size 選擇不當。若能透過 cross validation 調整超參數，並引入 Early Stopping 或 regularization 方法，應該可以縮小 differences。

4. 在處理表格型資料時，常見的做法大概可以分成三種：**Filter**（像是用 correlation 或 mutual information 來篩選特徵）、**Wrapper**（例如 Recursive Feature Elimination 會不斷嘗試不同的特徵組合來找出效果最好的）、還有 **Embedded**（像是 Lasso 在訓練模型的同時就會自動挑特徵）。透過這些方法做適當的特徵選擇，不只能讓模型變簡單、更容易理解，還能降低 overfitting 的風險，最終在新的資料上也比較容易有穩定、準確的表現。

Sources: Chandrashekar, G., & Sahin, F. (2014). *A survey on feature selection methods*. *Computers & Electrical Engineering*, 40(1), 16 – 28

5. 我選的模型是 TabNet。它的架構採用 sequential attention 的機制，特別針對異質性的特徵（heterogeneous features）去優化。跟傳統的人工神經網路（ANNs）不同，TabNet 會在每一步決策（decision step）中，透過 attentive feature selection 自動挑出對當前樣本最重要的變數來處理。這種設計對同時 numeric 和 categorical 資料的表格來說特別有優勢，因為這些資料不像影像或文字有明顯的空間或時間結構。另外，TabNet 採用的 sparse attention 機制能透過聚焦在真正關鍵的特徵來減少 overfitting。從實驗結果來看，TabNet 在處理大型的表格資料集時，表現甚至有時候可以比 gradient boosting 之類的傳統方法更好。

Sources: Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE transactions on neural networks and learning systems*