

# 2019 Spring COM526000 Deep Learning - Homework 2

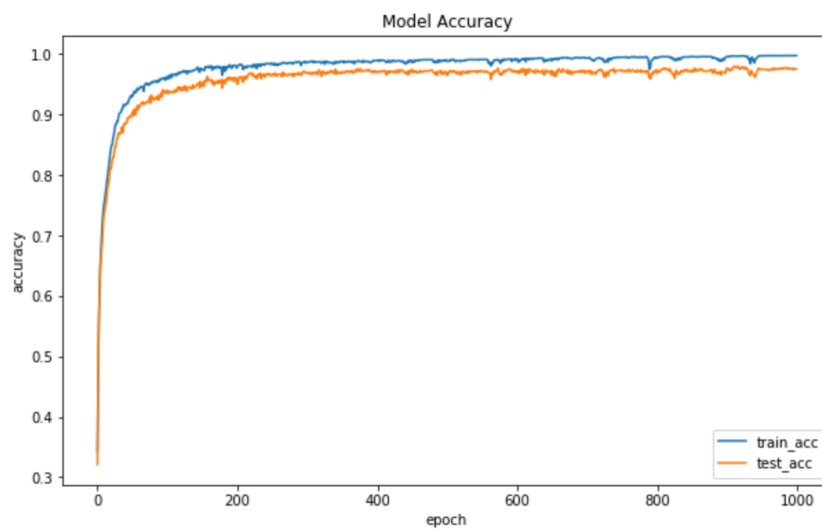
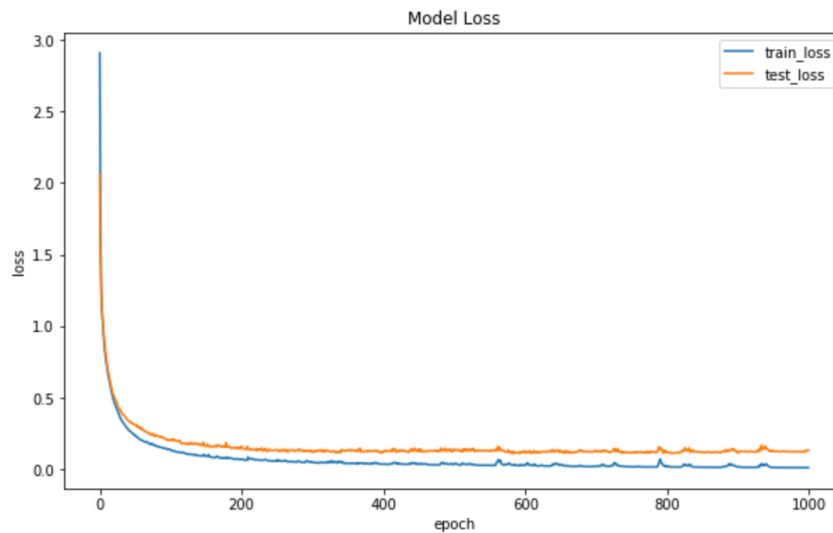
## Deep learning Model: Deep Neural Network

105061210 楊雅婷

### Problems

#### 1. Classification:

- i. DNN : number of hidden layers = 2, hidden unit = [60, 30], epochs = 1000, learning rate = 0.001, batch size = 128



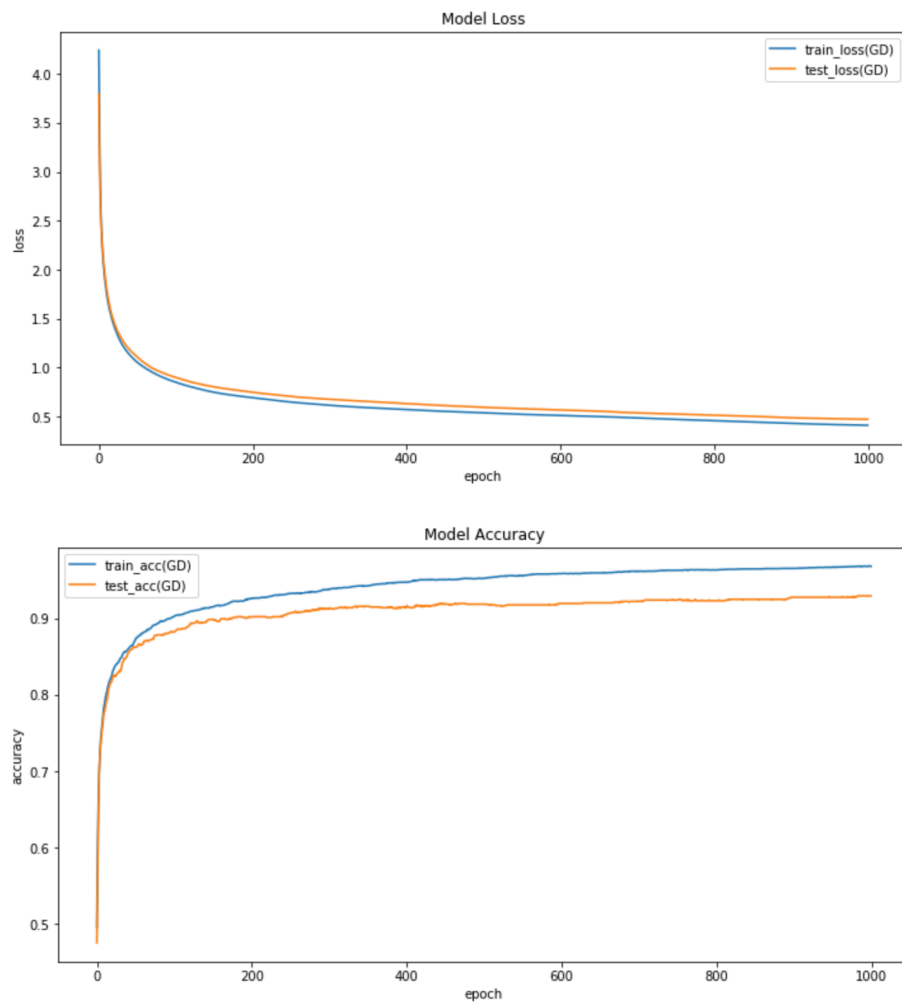
ii. Model performance :

DNN classification\_report:

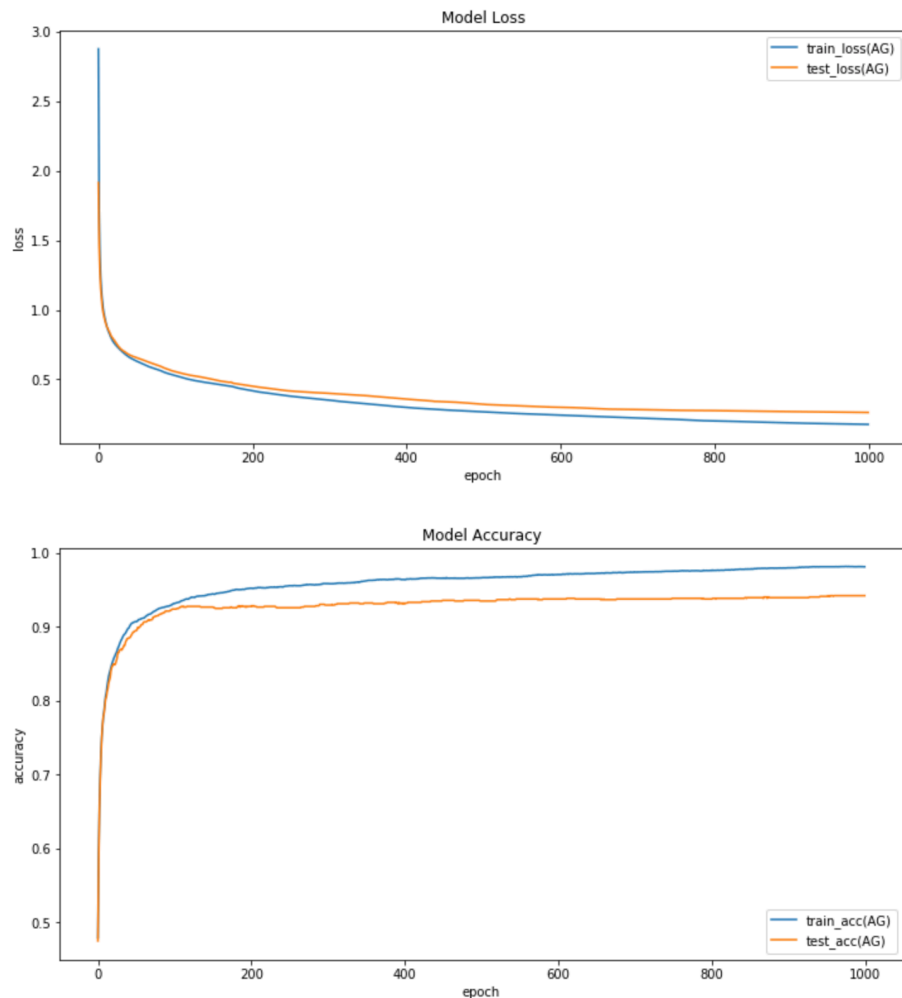
	precision	recall	f1-score	support
1	0.91	0.96	0.93	93
2	0.99	1.00	1.00	250
3	0.98	0.96	0.97	113
4	0.93	0.91	0.92	110
5	0.98	0.98	0.98	283
6	1.00	0.99	1.00	252
micro avg	0.98	0.98	0.98	1101
macro avg	0.97	0.97	0.97	1101
weighted avg	0.98	0.98	0.98	1101

iii. Different optimizers :

Gradient descent : learning rate = 0.001



Adagrad : learning rate = 0.01



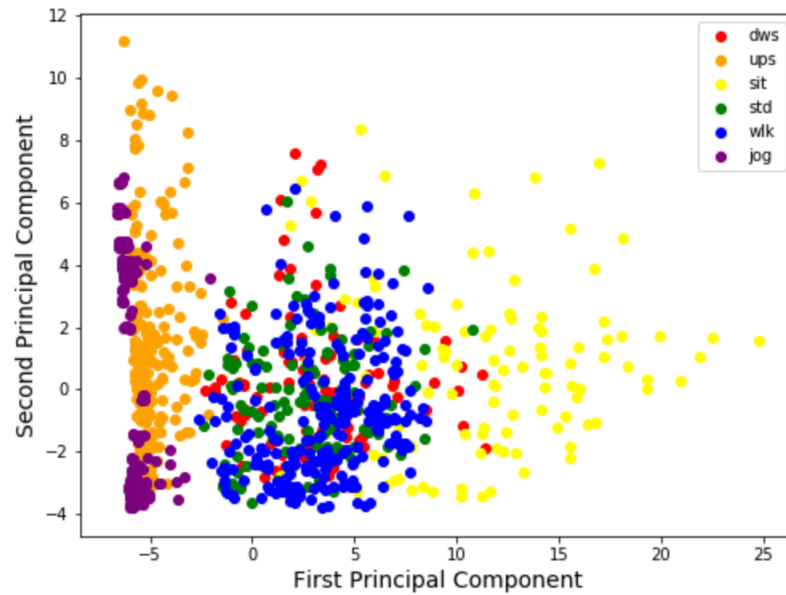
觀察調了幾次參數的結果下來：

loss : Gradient descent > Adagrad > Adam

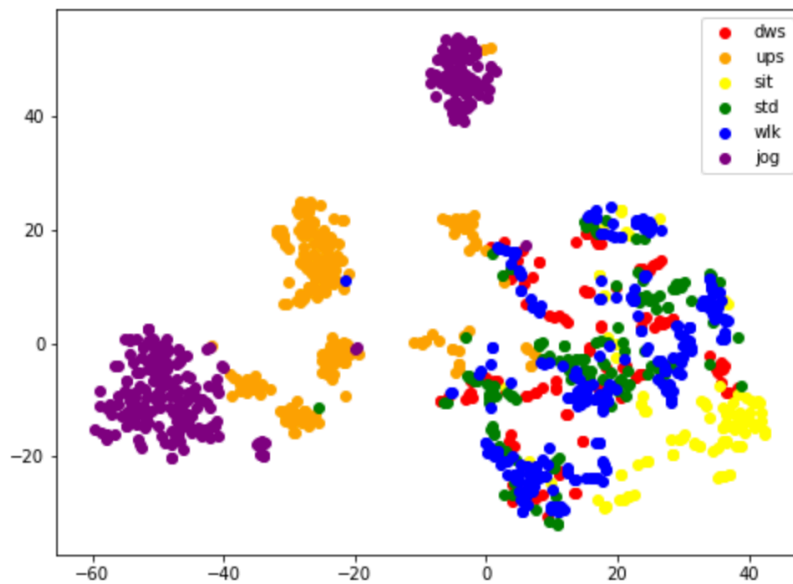
Accuracy : Adam > Adagrad > Gradient descent

Gradient descent 就是傳統上最直觀的optimizer方法，每次更新參數的 learning rate都是固定的；Adagrad則是讓learning rate隨著參數調整的幅度而改變（更新的總距離增多，學習速率也跟著變慢）；可以看出Adam的效果較好，這可能是因為Adam會依據計算梯度的一階估計和二階估計來為不同參數設計獨特的 learning rate。

iv. PCA :



v. t-SNE :



vi. 在PCA降維（數據標準化 -> 建立共變異數矩陣 -> SVD得到特徵值和特徵向量 -> 由大至小選k個特徵值和特徵向量 -> 將原本的數據投影到新的特徵向量上）後，由於這裡的 $k=2$ ，也就是只選擇了最重要的兩個特徵，所以我們只能保證最重要的兩個特徵相近的數據應該會聚集在一起，卻是直

接捨棄了其他可供判別的特徵，亦不能保證這兩個特徵不相近的數據會被分開；在t-SNE降維（高維數據用高斯分佈近似，低維數據用t分佈近似 -> 用KL距離算相似度 -> Gradient descent求最佳解）後，能使數據的分佈在高維和低維的空間中盡可能接近，保留了數據在高維度空間中的距離資訊，雖然要花較多時間，卻可以更好的維持數據的差異性。