機器學習 專案作業一

本作業主要使用Python語言及Keras/Tensorflow實作**前饋式神經網路預測類別與數值**。

作業內容包括：

* + 安裝與練習使用Anaconda裡的Jupyter Notebook系統。有問題可請教助教。
  + 資料集 (UCI ML Repository <https://archive.ics.uci.edu/ml/index.php>)  
    https://archive.ics.uci.edu/ml/datasets/Mushroom
  + 利用Python及Keras/Tensorflow建構前饋神經網路並分別進行以下操作。
    1. 第一個實驗資料集，請從UCI ML Repository中，自選一個具有數值與類別型態資料集進行實驗，筆數**至少三千筆**以上，並分別隨機將資料集分成：訓練資料70%，驗證資料15%，測試資料15%。請利用驗證資料選取最佳模型之參數，進行對測試資料之績效評估。
    2. 第一個實驗資料集中，請自選一組類別型欄位及一組數值型欄位並分別預測。
    3. 第二個實驗資料集，請從UCI ML Repository中，下載Adult Data Set，adult.data當作訓練資料集、adult.test當作測試資料集，並預測hours-per-week數值型欄位的數值。

https://archive.ics.uci.edu/ml/datasets/Adult

* + 1. 此資料集為混合型態，內含數值型欄位以及名目型欄位，名目型欄位值必須轉換成數值，才能進行模型訓練。
    2. 類別預測績效衡量請使用Precision, Recall及F1指標，數值預測使用MAE、MAPE與RMSE指標。
    3. 專案須包含下列要項：
    - 說明進行何種資料前置處理？
    - 說明類神經網路的架構與參數設定：number of hidden layers, number of units, activation function, optimizer, batch size, epoch, etc.
    - 請試著比較不同的架構造成什麼影響? (例如，隱藏層數量的變換、節點數量的變換 、batch size, epoch等，請自由發揮。)
    - 分別列出對訓練資料、驗證資料及測試資料的預測績效。
  + 請遵循學術論文寫作格式撰寫專案報告，內容至少必須包括下列要項：

**摘要** (學術論文摘要通常不分段，內容簡要地涵蓋動機、目的、方法與實驗結果)

**一、緒論**

**1.1動機** (選用的資料集之動機)

**1.2目的** (想進行哪種預測或分析，期望了解甚麼？)

**二、方法**

(1. 簡單說明程式架構，2. 執行程式的方法，無法執行程式，視為不正確)

**三、實驗**

**3.1資料集** (簡介資料集：名稱、筆數、欄位、型態…、並以資料表格呈現部份資料內容)

**3.2前置處理** (例如，名目型欄位值轉換成數字、欄位值正規化至0~1之間、圖像類別標記等)

**3.3實驗設計** (實驗如何進行、若有參數的話，參數如何設定等）

**3.4實驗結果** (呈現結果)

1. **結論**

* 評分標準:

內容(難度、豐富性、方法正確性)：60%，論文格式：20%，寫作技巧(文筆通順)：20%

* 任何複製剪貼網站資料或學長姐以前作業，皆屬抄襲，嚴重者零分。引用網站或他人資料，應註明引用來源。
* 此份報告最多15頁，請加註頁碼。請遵循投稿論文寫作格式，使用一欄式，無需製作論文目錄。可以觀摩 ACM Transactions on Knowledge Discovery from Data期刊論文的寫作技巧與格式。
* 紙本報告於上課前繳交或直接放入許老師的系辦信箱。專案程式直接上傳GitHub，專案名稱：hw1\_組長學號，並新增助教為專案成員，以利下載評分。缺紙本或程式皆視為未完成繳交作業。
* 繳交紙本時，在左上角訂一訂書針即可，無須其他任何裝訂。第一頁需有組員姓名及學號。
* 每遲交一天，扣該次作業成績10分。
* 繳交日期：112/04/13（四）