

# 機器學習 專案作業一

組員

M11221004 侯郡凌

M11223040 邱琳恩

中華民國一一三年四月十一日

## 摘要

本研究針對三個不同類型的資料集進行資料處理和神經網路建置，分別為 MNIST Data Set（用於圖像分類）、Boston Housing Data Set（用於預測房價）以及 Adult Data Set（用於收入分類和預測每週工時）。針對不同任務，本研究採用前饋式神經網路的方法進行神經網路建置，透過嘗試不同的超參數組合來尋找最佳的模型配置，以達到最優的預測效果，且為了解決過擬合問題，本研究採用 L2 正則化和 K-Fold 交叉驗證等技術。在 MNIST Data Set、Boston Housing Data Set，取得良好的績效表現，這表明模型在該資料集上能夠很好地泛化。然而，在 Adult Data Set，本研究認為還需要進一步嘗試不同的超參數組合，以提升模型的績效和泛化能力。

關鍵字：機器學習、分類、迴歸預測、前饋式神經網路、資料預處理

## 一、緒論

### 1.1 動機

本研究動機在於了解前饋式神經網路在三種不同類型的資料集上的應用與背後原理，神經網路作為一種強大機器學習模型，可透過模擬人腦神經元的方式透過多層的運算，實現更為複雜的任務，在圖像分類、迴歸預測和分類任務中取得了顯著的成果，為了解類神經網路模型背後的機制原理和面對不同類型的問題上之表現與限制，本研究選用 MNIST Data Set、Boston Housing Data Set、Adult Data Set 進行實作，透過處理圖像分類問題、房價預測問題、收入分類問題、工時預測問題來認識類神經網路的分類與迴歸任務進行方式，有助於更好地應用和優化這些模型，以解決現實中的各種問題。

### 1.2 目的

本研究目的在於以前饋式神經網路方式透過 MNIST Data Set、Boston Housing Data Set、Adult Data Set 進行實際的程式模擬，了解類神經網路的建置，以及前饋式神經網路在不同類型問題上的應用。

1. MNIST Data Set：透過此資料集，本研究欲了解神經網路在圖像分類任務上的表現，以及如何從圖片資料中學習和辨識數字。

2. Boston Housing Data Set：透過此資料集，本研究欲了解神經網路在迴歸預測任務上的能力，即在房屋價格預測中的表現，且期望了解神經網路是如何從各種相關特徵中學習，並預測房屋價格。

3. Adult Data Set：透過分析此資料集，本研究希望得知神經網路如何從每個人的資料中學習，並進行分類與預測任務。同時，本研究也將評估神經網路在處理複雜的分類任務上的泛化能力。

## 二、 方法

本研究的程式架構大致上可分為資料集的讀取、資料的前置處理、神經網路的建置、模型績效評估。

針對 MNIST Data set：首先透過 `keras.datasets` 的套件進行資料集的讀取，因 MNIST Data set 為手寫數字圖像資料集，所以針對圖像的每一個像素值進行操作，讓它可縮放到 0 和 1 之間。圖像分類任務在該資料集中屬於多類別單標籤問題，因此使用 `softmax` 為最後輸出的激活函數，最後以 Precision、Recall、F1 指標作為模型績效評估。

針對 Boston Housing Price Data Set：首先透過 `tensorflow.keras.datasets` 的套件進行資料集的讀取，經過觀察，因資料集已進行資料預處理，即無缺漏值且皆為數值型資料，因此，接著進行訓練集與測試集的特徵與目標切分，以訓練集的資料作為模型的輸入，使用測試集評估模型，且將資料進行正規化以及標準化，使模型不會因為資料尺度不同而受影響後，建立預測模型，並且因資料集較小較不穩定，因此加入 K-Fold 進行驗證與評估，最後選擇較合適的超參數進行訓練，最後以 MAE、MAPE、RMSE 指標作為模型績效評估。

針對 Adult Data Set：首先透過 `pandas` 套件從本地端進行資料的讀取，且因原資料集分為 `Adult.data` 和 `Adult.test`，在實驗中分別用於訓練資料和測試資料，並針對該資料集進行遺失值處理、移除無關欄位、資料型別轉換、正規化。本研究使用該資料集分別進行收入分類和預測每週工時的任務，為二元分類問題與迴歸預測問題，針對分類問題使用 `Sigmoid` 作為最後輸出的激活函數，迴歸預測問題則直接進行線性輸出，且為了應對過擬合問題，採用了 L2 正則化技術以及 K-Fold 技術。在模型績效評估部分，分類任務使用 Precision、Recall、F1 指標作為模型績效評估，迴歸預測部分使用 MAE、MAPE、RMSE 指標作為模型績效評估。

### 三、實驗

#### 3.1 資料集

本研究使用的資料集有：MNIST Data set、Boston Housing Data Set 以及 Adult Data Set。

MNIST Data set 資料說明：欄位包含 images(圖像)、labels(標籤)，資料型態皆為 uint8，訓練資料筆數為 6,000 筆資料，測試資料筆數為 1,000 筆資料，如表 1 所示。

Boston Housing Data Set 資料說明：訓練集共有 404 筆資料、測試集則有 102 筆資料，欄位包含 CRIM(各城鎮人均犯罪率)、ZN(劃分為 25,000 平方英尺以上地塊的住宅用地比例)、INDUS(每個城鎮非零售商業面積的比例。)、CHAS(查爾斯河虛擬變數，若區域邊界為 1，則為 1；否則為 0)、NOX(千萬分之一的一氧化氮濃度)、RM(每套住宅的平均房間數)、AGE(1940 年以前建造的自住單位的比例)、DIS(到五個波士頓就業中心的加權距離)、RAD(放射狀高速公路的可達性指數)、TAX(每 10,000 美元的全額財產稅稅率)、PTRATIO(各城鎮的師生比)、B( $1000(Bk-0.63)^2$ ，其中 Bk 是按城鎮劃分的黑人比例)、LSTAT(人口地位較低的百分比)、MEDV(自住房屋的中位數價格，約 1000 美元)，以及資料型態皆為 float64，部分資料呈現如表 2。

Adult Data Set 資料說明：資料欄位包含 age(年齡)、workclass(工作類別)、fnlwgt(序號)、education(教育程度)、education-num(受教育時間)、marital-status(婚姻狀況)、occupation(職業)、relationship(關係)、race(種族)、sex(性別)、capital-gain(資本收益)、capital-loss(資本支出)、hours-per-week(每週工時)、native-country(國籍)、income(收入)，其中 age(年齡)、education-num(受教育時間)、hours-per-week(每週工時)欄位為 int64 資料型態，其餘皆為 object 資料型態，訓練資料筆數為 32,561 筆資料，測試資料筆數為 16,281 筆資料，如表 3 所示。

表 1、MNIST Data set

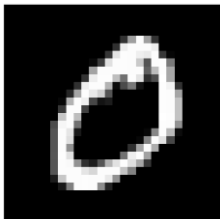
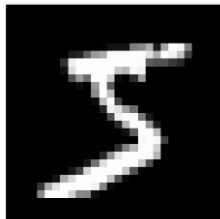
欄位名稱	資料型態	資料 1	資料 2
images	uint8		
labels	uint8	0	5

表 2 Boston Housing Data Set

欄位名稱	資料型態	資料 1	資料 2
CRIM	float64	0.00632	0.62739
ZN	float64	18	0
INDUS	float64	2.31	8.14
CHAS	float64	0	0
NOX	float64	0.538	0.538
RM	float64	6.575	5.834
AGE	float64	65.2	56.5
DIS	float64	4.09	4.4986
RAD	float64	1	4
TAX	float64	296	307
PTRATIO	float64	15.3	21
B	float64	396.9	395.62
LSTAT	float64	4.98	8.47
MEDV	float64	24	19.9

表 3 、Adult Data Set

欄位名稱	資料型態	資料 1	資料 2
age	int64	39	31
workclass	object	State-gov	Private
fnlwgt	object	77516	45781
education	object	Bachelors	Masters
education-num	int64	13	14
marital-status	object	Never-married	Never-married
occupation	object	Adm-clerical	Prof-specialty
relationship	object	Not-in-family	Not-in-family
race	object	White	White
sex	object	Male	Female
capital-gain	object	2174	14084
capital-loss	object	0	0
hours-per-week	int64	40	50
native-country	object	United-States	United-States
income	object	<=50K	>50K

## 3.2 前置處理

本研究針對三個不同的資料集進行資料前置處理，分別為 MNIST Data set、Boston Housing Data Set、Adult Data Set。

MNIST Data set：經過初步的分析，可觀察到此資料集無缺失值，可直接進行訓練集與測試集劃分，並針對圖像的像素值進行操作，透過將像素值除以 255 的方式讓每一像素值縮放到 0 到 1 之間。

Boston Housing Data Set：本研究經由觀察後，發現此資料集無缺失值，且資料皆為數值型態，因此，可直接進行訓練集與測試集劃分，完成資料預處理後進行特徵與目標切分，因資料集任務為預測收入，則目標欄位為 income，其餘則為特徵欄位，接著進行資料正規化與標準化。

Adult Data Set：經由初步分析後，可發現在 workclass、occupation、native-country 欄位有缺失值出現，且缺失值符號為「？」，本研究透過 Pandas 套件的 replace 方法將「？」取代為「NaN」，並透過 fillna 方法以及運用眾數的方式進行資料的填補。本研究透過 describe() 查看資料的敘述性統計，發現 fnlwgt 為人員序號，與本研究目標較無相關，且 capital-gain 和 capital-loss 欄位之標準差與其他欄位差距甚大且資料多數為 0，故將其移除，透過 unique() 檢查資料欄位唯一值時，發現訓練集與測試集中的 income 欄位的唯一值不相同，分別為「<=50K」 「>50K」 與 「<=50K.」 「>50K.」，因此進行統一處理皆改以「<=50K」 「>50K」形式呈現，且本研究透過 LabelEncoder 進行型態轉換和編碼動作，將 sex、workclass、education、marital-status、occupation、relationship、race、native-country、income 欄位從名目類型欄位轉成數值類型，接著，將訓練集與測試集進行特徵與目標切分後，針對兩資料集特徵欄位進行正規化與標準化。

## 3.3 實驗設計

實驗設計部分，在分類問題中透過 Precision、Recall、F1 指標評估在各個參數設定下模型的實驗結果，本研究會透過調整隱藏層數量、節點數量、batch size、epoch、激活函數，來觀察模型績效的變化，找出最佳的實驗結果。

在迴歸預測問題中，本研究使用不同超參數組合進行實驗，如調整隱藏層數、神經元個數、優化器、batch size 以及 epoch 等超參數，並且繪製出損失函數圖以觀察此模型在何時收斂，以及是否有過擬合情況，最後使用 MAE、MAPE 與 RMSE 作為評估模型的指標。

### 3.4 實驗結果

針對 MNIST Data set 分類任務:如表 4 所示，增加隱藏層的層數可增強模型的訓練，而對於 MNIST Data set 來說隱藏層節點數量為 1000/500 時可獲得更好的績效，但在 bath size 的部分考慮在其餘參數不變的狀況下，128 大小和 64 大小之間的差距不大但有訓練時間的差異，因此本研究認為可選用 64 bath size 作為參數設定，而在 epoch 設定部分，若將 epoch 設定為 8，可發現在其他條件不變的狀況下可以呈現更好的績效，因此本研究在 MNIST Data set 圖像分類問題實驗中，隱藏層數量為 2、隱藏層節點為 1000/500、隱藏層和最後層激活函數分別為 ReLU/Softmax，優化器為 RMSprop、loss function 為 categorical cross entropy、bath size 為 64、epoch 為 8 可得到最好的績效表現。

針對 Boston Housing Data Set 預測任務：本研究透過 K-Fold 驗證發現，在 epoch 在 100-200 之間，模型開始趨於穩定。經過多次試驗和時間考量後，最終確定將 epoch 設定為 130。根據表 5 的結果，首先將節點數設為 64/64，取得了良好的表現，隨後將第一層節點數增加至 128 個、batch size 設為 64，測試後發現績效下降，從結果得知較大的 batch size 不適合較小的節點數；然而，當隱藏層數量增加至 3 層並增加節點數時，模型績效有所提升，最後將 batch size 設為 16，發現績效反而下降。因此，本研究認為在此模型中，最佳超參數組合為：隱藏層數為 3、節點數為 128/64/64、激活函數皆為 Relu、最佳化器為 rmsprop、batch size 為 32、epoch 為 130，在 Boston Housing Data Set 上可以得到最佳的績效表現。

針對 Adult Data Set 分類任務：如表 6 所示，本研究嘗試將隱藏層的激活函數調為 Sigmoid，但在其他參數不變的狀況下可發現在隱藏層使用 Sigmoid 會造成績效下降，表示在該模型當中不適合使用 Sigmoid 作為隱藏層的激活函數。且在 Adult Data Set 分類問題實驗的結果中可發現，其他參數不變下，bath size 調整為 128 大小，會讓績效下降，本研究推測結果是模型過度訓練導致的過擬合，在 epoch 設定部分，可發現若將 epoch 設定為 8，可得到比 epoch 設定為 5 更好的績效表現，因此本研究在 Adult Data Set 分類問題實驗中，當隱藏層數量為 3、隱藏層節點為 500/250/125、隱藏層和最後層激活函數分別為 ReLU/Sigmoid，優化器為 RMSprop、loss function 為 binary cross entropy、bath size 為 64、epoch 為 8 可得到最好的績效。

針對 Adult Data Set 預測任務：如表 7 所示，在考慮隱藏層數量、節點數、激活函數、優化器等方面的變化後，本研究發現增加隱藏層數量和節點數並不總是能夠提高模型績效。例如：將節點數設定為 500/500 和 1024/1024 時模型績效卻未提升；相反，較小的節點數可能會產生更好的表現，而較大的節點數會增加訓練時間且無法提升模型績效，而使用 Relu 激活函數和 rmsprop 優化器的組合通常能夠獲得較好的效果。當節點數為 64/64 且 batch size 為 32 時，Epoch 則是設為 10，可以讓模型績效稍微提升，不同的 batch size 和 epoch 組合可能會

導致績效有些微變化。

綜合結果來說，以 Adult Data Set 的分類和預測模型，本研究認為需要再進行更多的實驗和調整來找到最佳的超參數組合，以達到最佳的模型績效。

表 4、MNIST Data set 分類實驗結果

number of hidden layers	number of units	activation function (hidden layer)	optimizer	loss function	batch size	epoch	Train			Test		
							precision	recall	f1_score	precision	recall	f1_score
1	1000	ReLU/Softmax	RMSprop	categorical cross_entropy	128	5	0.992	0.990	0.991	0.982	0.984	0.981
1	1000	Sigmoid /Softmax	RMSprop	categorical cross_entropy	128	5	0.965	0.948	0.956	0.957	0.965	0.951
2	1000/500	ReLU/Softmax	RMSprop	categorical cross_entropy	128	5	0.993	0.992	0.992	0.984	0.985	0.983
1	500	ReLU/Softmax	RMSprop	categorical cross_entropy	128	5	0.990	0.987	0.989	0.977	0.980	0.976
<b>2</b>	<b>1000/500</b>	<b>ReLU/Softmax</b>	<b>RMSprop</b>	<b>categorical cross_entropy</b>	<b>64</b>	<b>8</b>	<b>0.997</b>	<b>0.997</b>	<b>0.997</b>	<b>0.983</b>	<b>0.983</b>	<b>0.983</b>
2	1000/500	ReLU/Softmax	RMSprop	categorical cross_entropy	128	8	0.997	0.997	0.997	0.984	0.984	0.984

表 5 Boston Housing Data Set 預測實驗結果

number of hidden layers	number of units	activation function (hidden layer)	optimizer	loss function	batch size	epoch	Train			Test		
							MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE
2	64/64	relu/relu	rmsprop	mse	16	130	1.51	7.58	2.11	2.84	17.06	4.28
2	128/64	relu/relu	rmsprop	mse	64	130	1.84	9.11	2.50	2.50	16.64	4.27
<b>3</b>	<b>64/64/32</b>	<b>relu/relu/relu</b>	<b>rmsprop</b>	<b>mse</b>	<b>16</b>	<b>130</b>	<b>1.40</b>	<b>7.91</b>	<b>1.92</b>	<b>2.60</b>	<b>16.91</b>	<b>4.00</b>
<b>3</b>	<b>64/64/32</b>	<b>relu/relu/relu</b>	<b>rmsprop</b>	<b>mse</b>	<b>32</b>	<b>130</b>	<b>1.41</b>	<b>7.04</b>	<b>1.86</b>	<b>2.45</b>	<b>14.44</b>	<b>3.60</b>
3	64/64/32	relu/relu/relu	rmsprop	mse	64	130	1.54	7.58	2.21	2.48	14.52	3.84
3	128/64/32	relu/relu/relu	rmsprop	mse	16	130	1.46	6.91	2.05	2.56	14.04	3.52
<b>3</b>	<b>128/64/32</b>	<b>relu/relu/relu</b>	<b>rmsprop</b>	<b>mse</b>	<b>32</b>	<b>130</b>	<b>1.20</b>	<b>6.32</b>	<b>1.63</b>	<b>2.51</b>	<b>15.91</b>	<b>4.09</b>
<b>3</b>	<b>128/64/64</b>	<b>relu/relu/relu</b>	<b>rmsprop</b>	<b>mse</b>	<b>32</b>	<b>130</b>	<b>1.18</b>	<b>6.08</b>	<b>1.67</b>	<b>2.40</b>	<b>14.02</b>	<b>3.62</b>
3	128/128/64	relu/relu/relu	rmsprop	mse	16	130	1.71	9.27	2.04	2.62	15.71	3.60



表 6、Adult Data Set 分類實驗結果

number of hidden layers	number of units	activation function (hidden layer)	optimizer	loss function	batch size	epoch	Train			Test		
							precision	recall	f1_score	precision	recall	f1_score
1	1000	ReLU/Sigmoid	RMSprop	binary cross_entropy	64	5	0.301	0.705	0.204	0.284	0.667	0.192
1	500	ReLU/Sigmoid	RMSprop	binary cross_entropy	64	5	0.575	0.681	0.526	0.569	0.674	0.519
1	500	ReLU/Sigmoid	RMSprop	binary cross_entropy	128	5	0.555	0.681	0.497	0.549	0.678	0.490
1	500	ReLU/Sigmoid	RMSprop	binary cross_entropy	64	8	0.573	0.685	0.526	0.569	0.679	0.520
1	500	Sigmoid/Sigmoid	RMSprop	binary cross_entropy	64	8	0.088	0.324	0.053	0.081	0.292	0.048
2	500/250	ReLU/Sigmoid	RMSprop	binary cross_entropy	64	8	0.514	0.723	0.425	0.509	0.717	0.420
<b>3</b>	<b>500/250/125</b>	<b>ReLU/Sigmoid</b>	<b>RMSprop</b>	<b>binary cross_entropy</b>	<b>64</b>	<b>8</b>	<b>0.639</b>	<b>0.528</b>	<b>0.848</b>	<b>0.633</b>	<b>0.524</b>	<b>0.841</b>
3	1500/1000/500	ReLU/Sigmoid	RMSprop	binary cross_entropy	64	5	0.613	0.628	0.634	0.607	0.617	0.628

表 7 Adult Data Set 預測實驗結果

number of hidden layers	number of units	activation function (hidden layer)	optimizer	loss function	batch size	epoch	Train			Test		
							MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE
2	64/32	Relu/Relu	rmsprop	MSE	16	5	7.66	31.54	11.01	7.72	31.78	11.08
2	64/64	Relu/Relu	rmsprop	MSE	32	5	7.69	32.45	11.10	7.74	32.65	11.18
<b>2</b>	<b>64/64</b>	<b>Relu/Relu</b>	<b>rmsprop</b>	<b>MSE</b>	<b>32</b>	<b>10</b>	<b>7.46</b>	<b>31.03</b>	<b>10.79</b>	<b>7.55</b>	<b>31.50</b>	<b>10.93</b>
2	128/64	Relu/Relu	rmsprop	MSE	16	10	7.56	30.16	10.92	7.64	30.55	11.04
2	128/64	Relu/Relu	rmsprop	MSE	16	5	7.72	31.42	11.09	7.78	31.74	11.19
2	128/64	Relu/Relu	rmsprop	MSE	32	5	7.69	32.24	11.06	7.74	32.52	11.15
2	128/64	Relu/Relu	rmsprop	MSE	32	10	7.59	30.35	10.92	7.65	30.78	11.02
2	128/128	Relu/Relu	rmsprop	MSE	32	10	7.57	30.18	10.87	7.65	30.64	11.01
3	128/128/64	Relu/Relu	rmsprop	MSE	32	10	7.76	30.42	11.10	7.89	30.96	11.27
2	500/500	Relu/Relu	rmsprop	MSE	32	10	7.61	30.88	10.92	7.77	31.54	11.13
2	1024/1024	Relu/Relu	rmsprop	MSE	64	10	7.83	32.33	11.07	7.89	32.66	11.18

#### 四、 結論

本研究針對三個資料集皆進行了資料預處理、資料編碼和模型訓練等步驟。對於 MNIST Data set，透過觀察後發現資料沒有遺失值，因此進行訓練集與測試集的特徵與目標切分，最後選擇適合的超參數組合進行訓練和評估。對於 Boston Housing Data Set，資料已進行預處理，因此直接進行訓練集與測試集的特徵與目標切分，並將資料正規化和標準化，也採用 K-Fold 進行了模型驗證與評估，並最終選擇合適的超參數組合進行了訓練和評估。對於 Adult Data Set，先對資料進行缺失值處理，使用眾數填充缺失值並移除較無關的欄位，接著進行資料正規化與標準化處理且採用了 L2 正則化技術和 K-Fold 驗證來應對過擬合問題，並測試了不同超參數組合進行訓練和評估。

透過本研究可得知，激活函數選擇以及適當超參數設定對於前饋式神經網路實驗的重要性，在一般狀況之下使用 Sigmoid 作為隱藏層激活函數很有可能導致績效下降。而在輸出層的激活函數選擇必須要根據任務類型進行調整，若是二元分類問題採用 Sigmoid，多類別單標籤問題採用 Softmax，而在迴歸預測問題上不選用激活函數，直接進行輸出較為恰當。進行實驗的同時也須注意模型的過擬合問題發生，使用 L2 正則化和 K-Fold 交叉驗證技術可有效提升模型穩定度和績效。

綜合實驗結果發現，在 MNIST Data Set、Boston Housing Data Set 中，採用適當的參數設定能夠在績效上取得較好的表現，而在 Adult Data Set 中，本研究認為仍需要進一步的實驗以調整模型與超參數以達到最佳模型績效。

## 參考文獻

10 程式中. (2020, September 20). [Day 5] 資料清理&前處理. It 邦幫忙.

<https://ithelp.ithome.com.tw/articles/10240494>

10 程式中. (2020, September 20). Python 全民瘋 AI 系列 [Day 5] 資料

清理&前處理 [Video]. YouTube.

<https://www.youtube.com/watch?v=P42GqxCXkY8>

10 程式中. (2020, September 20). [Day 5] 資料清理&前處理. It 邦幫忙.

<https://ithelp.ithome.com.tw/articles/10240494>

Evan. (2020, February 11). 零基礎自學深度學習 : (二) 優化神經網路.

Medium.

<https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4>

[%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5](https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5)

[%AD%B8%E7%BF%92-%E4%BA%8C-](https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5)

[%E5%84%AA%E5%8C%96%E7%A5%9E%E7%B6%93%E7%B6%](https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5)

[B2%E8%B7%AF-d2c71b5df7e5](https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5)

Evan. (2020, February 11). 零基礎自學深度學習 : (二) 優化神經網路.

Medium.

<https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4>

[%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5](https://evanhsiao.medium.com/%E9%9B%B6%E5%9F%BA%E7%A4%E8%87%AA%E5%AD%B8%E6%B7%B1%E5%BA%A6%E5)

%AD%B8%E7%BF%92-%E4%BA%8C-

%E5%84%AA%E5%8C%96%E7%A5%9E%E7%B6%93%E7%B6%

B2%E8%B7%AF-d2c71b5df7e5

iThome. (n.d.). DAY10 資料前處理-資料編碼、資料切割 - IT 邦幫忙::

一起幫忙解決難題,拯救 IT 人的一天. iT 邦幫忙::一起幫忙解決難  
題,拯救 IT 人的一天.

<https://ithelp.ithome.com.tw/articles/10260683>

iThome. (n.d.-a). 【Day 4】資料預處理 Data Preprocessing ( 一 ) - IT

邦幫忙::一起幫忙解決難題,拯救 IT 人的一天. iT 邦幫忙::一起幫  
忙解決難題,拯救 IT 人的一天.

<https://ithelp.ithome.com.tw/articles/10322270>

iThome. (n.d.-a). 【Day 5】資料預處理 Data Preprocessing ( 二 ) - IT

邦幫忙::一起幫忙解決難題,拯救 IT 人的一天. iT 邦幫忙::一起幫  
忙解決難題,拯救 IT 人的一天.

<https://ithelp.ithome.com.tw/articles/10322489>

James, Y. (2022, July 25). [資料分析&機器學習] 第 2.4 講：資料前處理

(Missing data, One-hot encoding, Feature Scaling). Medium.

[https://medium.com/jameslearningnote/%E8%B3%87%E6%96%99%](https://medium.com/jameslearningnote/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90%E6%A9%9F%E5%99%A8%E5%AD%B8)

[E5%88%86%E6%9E%90%E6%A9%9F%E5%99%A8%E5%AD%B8](https://medium.com/jameslearningnote/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90%E6%A9%9F%E5%99%A8%E5%AD%B8)

%E7%BF%92-%E7%AC%AC2-4%E8%AC%9B-

%E8%B3%87%E6%96%99%E5%89%8D%E8%99%95%E7%90%86

-missing-data-one-hot-encoding-feature-scaling-3b70a7839b4a

Wang, S. (2020, May 15). 機器學習模型評估指標-Confusion Matrix,

Precision, Recall, and ,F1-Score. Medium.

<https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A>

[D%B8%E7%BF%92%E6%A8%A1%E5%9E%8B%E8%A9%95%E4](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)

[%BC%B0%E6%8C%87%E6%A8%99-confusion-matrix-precision-](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)

[and-recall-e9d64ff14d81](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)

Wang, S. (2020, May 15). 機器學習模型評估指標-Confusion Matrix,

Precision, Recall, and ,F1-Score. Medium.

<https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A>

[D%B8%E7%BF%92%E6%A8%A1%E5%9E%8B%E8%A9%95%E4](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)

[%BC%B0%E6%8C%87%E6%A8%99-confusion-matrix-precision-](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)

[and-recall-e9d64ff14d81](https://medium.com/@s716419/%E6%A9%9F%E5%99%A8%E5%A)