深度學習導論與應用
作業1 - ANN Classification

第9組 - 409411167林嘉柔
410402226林妍伶

目錄

變數定義

Group9

1047筆資料

8個變項

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked	Survived
0	3	male	NaN	0	0	7.7500	Q	no
1	3	female	NaN	0	0	8.0500	S	no
2	2	female	57.0	0	0	10.5000	S	no
3	3	male	NaN	0	0	7.2500	S	no
4	2	male	25.0	0	0	13.0000	S	no
...	...	...	...	...	...	...	...	...
1042	1	female	19.0	0	2	26.2833	S	yes
1043	3	male	25.0	0	0	7.7958	S	no
1044	2	female	29.0	1	0	26.0000	S	yes
1045	2	male	29.0	1	0	27.7208	C	no
1046	3	male	NaN	1	0	19.9667	S	no

1047 rows × 8 columns

變數定義

	變數	定義	類別值	資料類型
X1	Pclass	船艙等級	1 / 2 / 3	類別 (Category)
X2	Sex	性別	male / female	類別 (Category)
X3	Age	年齡		連續 (Numeric)
X4	SibSp	乘客在船上的兄弟姊妹 / 配偶人數		連續 (Numeric)
X5	Parch	乘客在船上的兄弟船上父母 / 子女人數		連續 (Numeric)
X6	Fare	乘客票價		連續 (Numeric)
X7	Embarked	登船港	C = Cherbourg Q = Queenstown S = Southampton	類別 (Category)
Y	Survived	是否生存	yes / no	類別 (Category)

敘述統計

敘述統計 - 連續型變項

變項	平均數	標準差	min	Q1 (25%)	Q2 (50%)	Q3 (75%)	max
Age	30.27	14.36	1	21	28	39	80
SibSp	0.50	1.04	0	0	0	1	8
Parch	0.40	0.90	0	0	0	0	9
Fare	33.27	50.98	0	0.79	14.46	31.39	512.33

敘述統計 - 類別型變項

變項	值	資料筆數	百分比
Pclass	1	264	25.21%
	2	220	21.01%
	3	563	53.77%

變項	值	資料筆數	百分比
Sex	male	679	64.85%
	female	368	35.15%
Embarked	C	223	21.30%
	Q	95	9.07%
	S	729	69.63%
Survived	yes	382	36.49%
	no	665	63.51%

資料前處理

資料前處理

資料筆數 (rows) : 1,047 → 799

欄位數 (columns) : 8 → 13

刪除有缺失值的資料

若有缺失值，則刪去整筆資料

類別變項轉為dummy variable

- Pclass : 1 / 2 / 3
- Sex : male / female
- Embarked : C / Q / S
- Survived : yes / no

設定自變數x、依變數y

- x : Pclass, Sex, Age, SibSp, Parch, Fare, Embarked
- y : Survived

敘述統計 - 連續型變項

before（1047筆資料）							
變項	平均數	標準差	min	Q1 (25%)	Q2 (50%)	Q3 (75%)	max
Age	30.27	14.36	1	21	28	39	80
SibSp	0.50	1.04	0	0	0	1	8
Parch	0.40	0.90	0	0	0	0	9
Fare	33.27	50.98	0	0.79	14.46	31.39	512.33

敘述統計 - 連續型變項

after（799筆資料）							
變項	平均數	標準差	min	Q1 (25%)	Q2 (50%)	Q3 (75%)	max
Age	30.27	14.36	1	21	28	39	80
SibSp	0.50	0.90	0	0	0	1	5
Parch	0.43	0.86	0	0	0	1	6
Fare	37.16	55.55	0	8.05	15.85	39.0	512.33

敘述統計 - 類別型變項

before (1047筆資料)			
變項	**值**	資料筆數	百分比
Pclass	1	264	25.21%
	2	220	21.01%
	3	563	53.77%

before (1047筆資料)			
變項	**值**	**資料筆數**	**百分比**
Sex	male	679	64.85%
	female	368	35.15%
Embarked	C	223	21.30%
	Q	95	9.07%
	S	729	69.63%
Survived	yes	382	36.49%
	no	665	63.51%

敘述統計 - 類別型變項

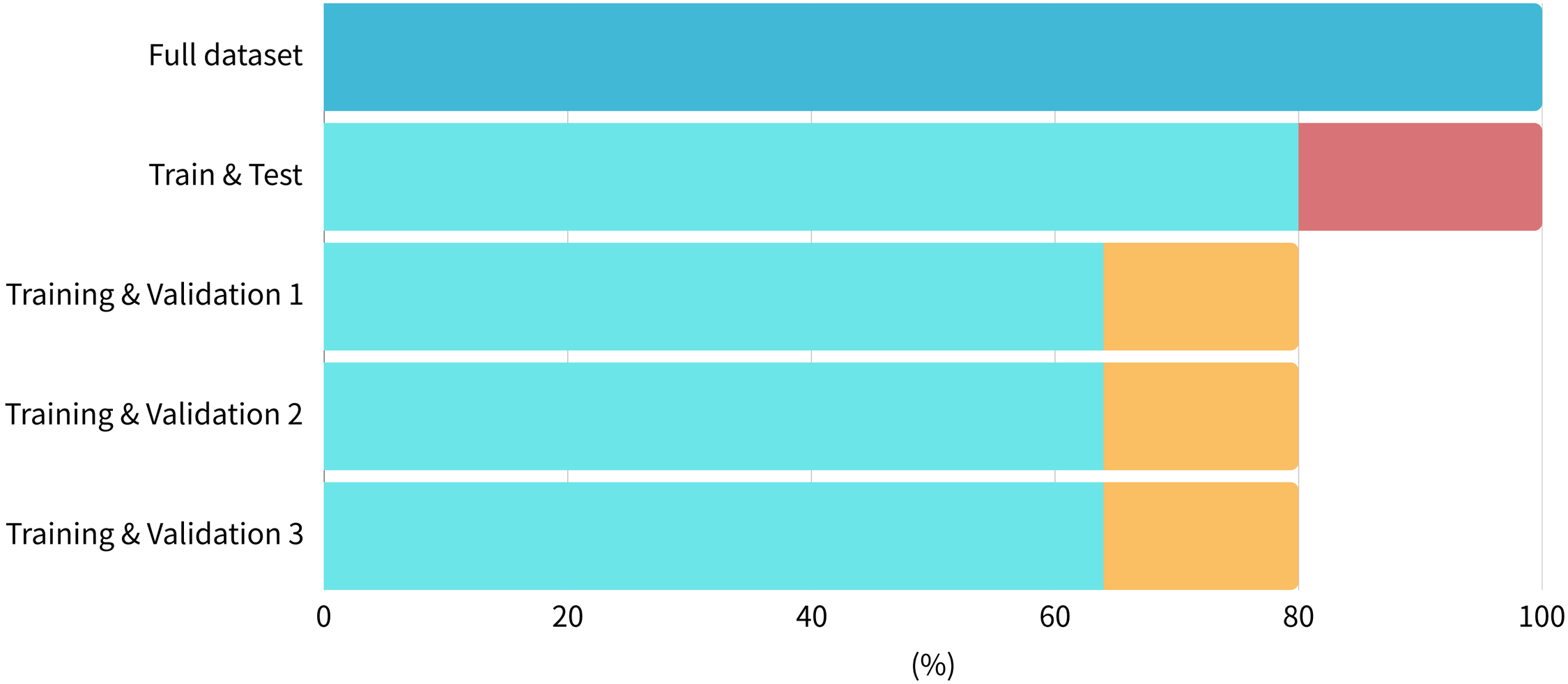
after (799筆資料)			
變項	**值**	資料筆數	百分比
Pclass	1	229	28.66%
	2	200	25.03%
	3	370	46.31%

after (799筆資料)			
變項	**值**	**資料筆數**	**百分比**
Sex	male	500	62.58%
	female	299	37.42%
Embarked	C	161	20.15%
	Q	37	4.63%
	S	601	75.22%
Survived	yes	311	38.92%
	no	488	61.08%

資料建模與結果

資料切分

holdout + holdout method



資料特徵縮放

針對x_train做scaling，把每一個特徵的數值壓縮到 0 到 1 的範圍內。
再根據得到的值(現有資料)來對x_test做轉換。

	Age	SibSp	Parch	Fare	Pclass_1	Pclass_2	Pclass_3	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S
0	0.586667	0.0	0.000000	0.015127	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
1	0.240000	0.0	0.000000	0.015412	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
2	0.640000	0.2	0.166667	0.216430	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
3	0.360000	0.2	0.000000	0.160387	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
4	0.320000	0.0	0.000000	0.014932	0.0	0.0						
...	...	...	...	...	...	...						
634	0.653333	0.4	0.000000	0.260867	1.0	0.0						
635	0.440000	0.0	0.000000	0.025374	0.0	1.0						
636	0.293333	0.0	0.000000	0.026920	0.0	1.0						
637	0.320000	0.0	0.000000	0.014932	0.0	0.0						
638	0.386667	0.0	0.000000	0.025374	0.0	1.0						

	Age	SibSp	Parch	Fare	Pclass_1	Pclass_2	Pclass_3	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S
0	0.306667	0.0	0.000000	0.015713	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
1	0.213333	0.2	0.166667	0.014110	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0
2	0.440000	0.0	0.000000	0.025374	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0
3	0.360000	0.0	0.000000	0.015176	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
4	0.506667	0.2	0.166667	0.162314	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...
155	0.253333	0.0	0.000000	0.007832	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0
156	0.400000	0.0	0.000000	0.262527	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
157	0.573333	0.4	0.000000	0.175668	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0
158	0.613333	0.2	0.000000	0.028302	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
159	0.280000	0.2	0.333333	0.081157	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0

固定超參數

- Hidden layers = 2
- Batch_size = 64
- Epoch = 100
- Activation function:
  - Hidden = relu
  - Output = sigmoid

改變的超參數組合

- random seed : 1 / 50 / 999
- learning rate : 0.001 / 0.05/ 0.1
- momentum : 0.5 / 0.7 / 0.9

超參數組合

random seed = 1			
	momentum = 0.5	**momentum = 0.7**	**momentum = 0.9**
learning rate = 0.001	val_loss = 0.6183 val_accuracy = 0.6719 RMSE = 0.4623	val_loss = 0.5985 val_accuracy = 0.6875 RMSE = 0.4522	val_loss = 0.5679 val_accuracy = 0.7266 RMSE = 0.4373
learning rate = 0.05	val_loss = 0.5460 val_accuracy = 0.7422 RMSE = 0.4261	val_loss = 0.5400 val_accuracy = 0.7344 RMSE = 0.4247	val_loss = 0.5625 val_accuracy = 0.7422 RMSE = 0.4321
learning rate = 0.1	val_loss = 0.5428 val_accuracy = 0.7266 RMSE = 0.4255	val_loss = 0.5520 val_accuracy = 0.7344 RMSE = 0.4299	val_loss = 0.5879 val_accuracy = 0.7422 RMSE = 0.4436

超參數組合

random seed = 50			
	momentum = 0.5	**momentum = 0.7**	**momentum = 0.9**
learning rate = 0.001	val_loss = 0.5919 val_accuracy = 0.7031 RMSE = 0.4484	val_loss = 0.5767 val_accuracy = 0.6953 RMSE = 0.4416	val_loss = 0.5038 val_accuracy = 0.7969 RMSE = 0.4021
learning rate = 0.05	val_loss = 0.5002 val_accuracy = 0.7734 RMSE = 0.4005	val_loss = 0.5317 val_accuracy = 0.7812 RMSE = 0.4105	val_loss = 0.5485 val_accuracy = 0.7891 RMSE = 0.4156
learning rate = 0.1	val_loss = 0.5363 val_accuracy = 0.8047 RMSE = 0.4046	val_loss = 0.5470 val_accuracy = 0.7969 RMSE = 0.4084	val_loss = 0.5762 val_accuracy = 0.7734 RMSE = 0.4221

超參數組合

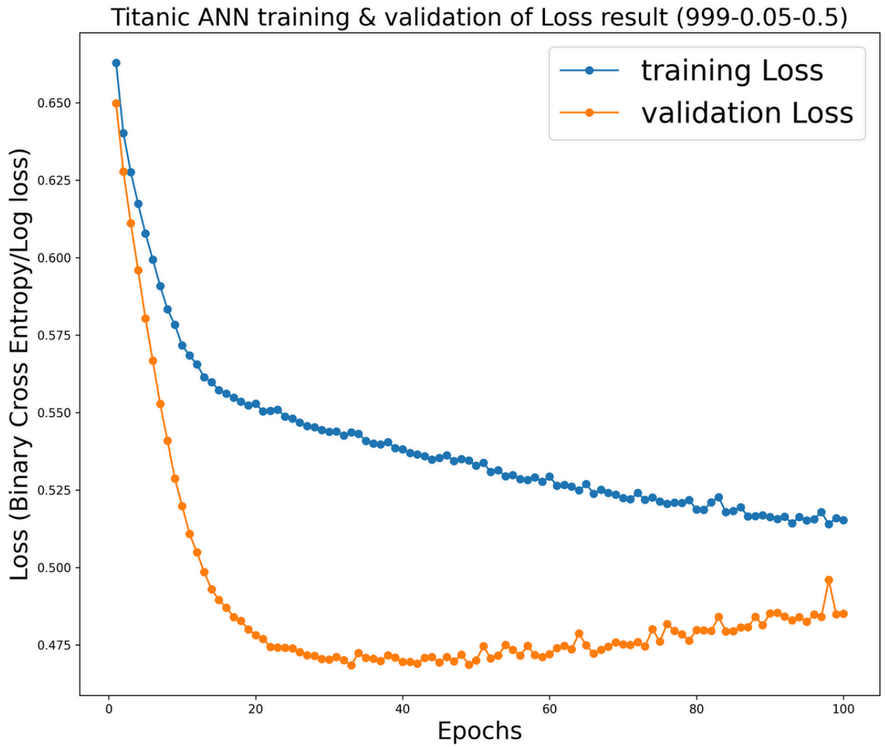
random seed = 999			
	momentum = 0.5	**momentum = 0.7**	**momentum = 0.9**
learning rate = 0.001	val_loss = 0.6116 val_accuracy = 0.6484 RMSE = 0.4585	val_loss = 0.6055 val_accuracy = 0.6172 RMSE = 0.4563	val_loss = 0.5134 val_accuracy = 0.8125 RMSE = 0.4075
learning rate = 0.05	val_loss = 0.4852 val_accuracy = 0.8125 RMSE = 0.3936	val_loss = 0.4906 val_accuracy = 0.7891 RMSE = 0.3962	val_loss = 0.5276 val_accuracy = 0.8125 RMSE = 0.4027
learning rate = 0.1	val_loss = 0.4923 val_accuracy = 0.8203 RMSE = 0.3946	val_loss = 0.5112 val_accuracy = 0.7969 RMSE = 0.4051	val_loss = 0.5166 val_accuracy = 0.8047 RMSE = 0.405

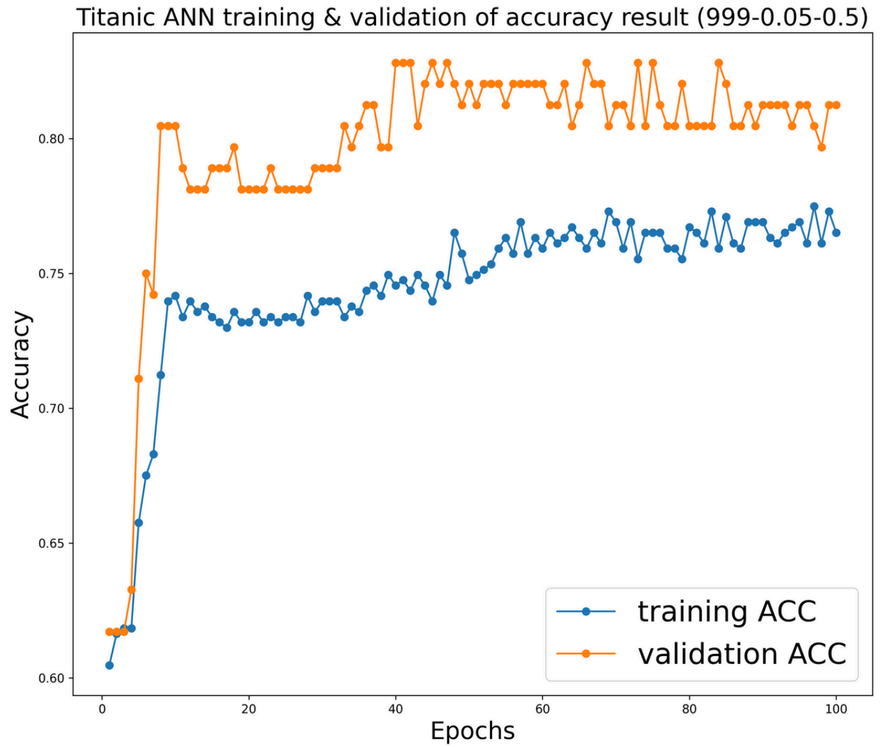
超參數組合 (跨random seed)

Average			
	momentum = 0.5	**momentum = 0.7**	**momentum = 0.9**
learning rate = 0.001	val_loss = 0.6073 val_accuracy = 0.6745 RMSE = 0.4564	val_loss = 0.5936 val_accuracy = 0.6667 RMSE = 0.4500	val_loss = 0.5284 val_accuracy = 0.7787 RMSE = 0.4156
learning rate = 0.05	val_loss = 0.5105 val_accuracy = 0.7760 RMSE = 0.4067	val_loss = 0.5208 val_accuracy = 0.7682 RMSE = 0.4105	val_loss = 0.5462 val_accuracy = 0.7813 RMSE = 0.4168
learning rate = 0.1	val_loss = 0.5238 val_accuracy = 0.7839 RMSE = 0.4082	val_loss = 0.5367 val_accuracy = 0.7761 RMSE = 0.4145	val_loss = 0.5602 val_accuracy = 0.7734 RMSE = 0.4236

Training & Validation Loss / Accuracy

random seed = 999, learning rate = 0.05, momentum = 0.5





測試結果

- 選用的超參數組合：
  - random seed = 999
  - learning rate = 0.05
  - momentum = 0.5
- Best thresholds = 0.36357152
- Testing result
  - RMSE : 0.4023
  - f1s (F-score) : 76.81
  - AUC(Area Under the Curve) : 81.38
  - ACC (Accuracy) : 80.0
  - pre (Precision) : 74.65
  - sen (Sensitivity) : 79.1
  - spe (Specificity) : 80.65

Performance Metrics - Testing

	Predicted : Positive	**Predicted : Negative**	
Actual : Positive	(TP) 53	(FN) 14	Sensitivity (sen) 79.1
Actual : Negative	(FP) 18	(TN) 75	Specificity (spe) 80.65
	Precision (pre) 74.65		Accuracy (ACC) 80.0

END