



分類模型評估方法



國立宜蘭大學資訊工程系
吳政瑋助理教授

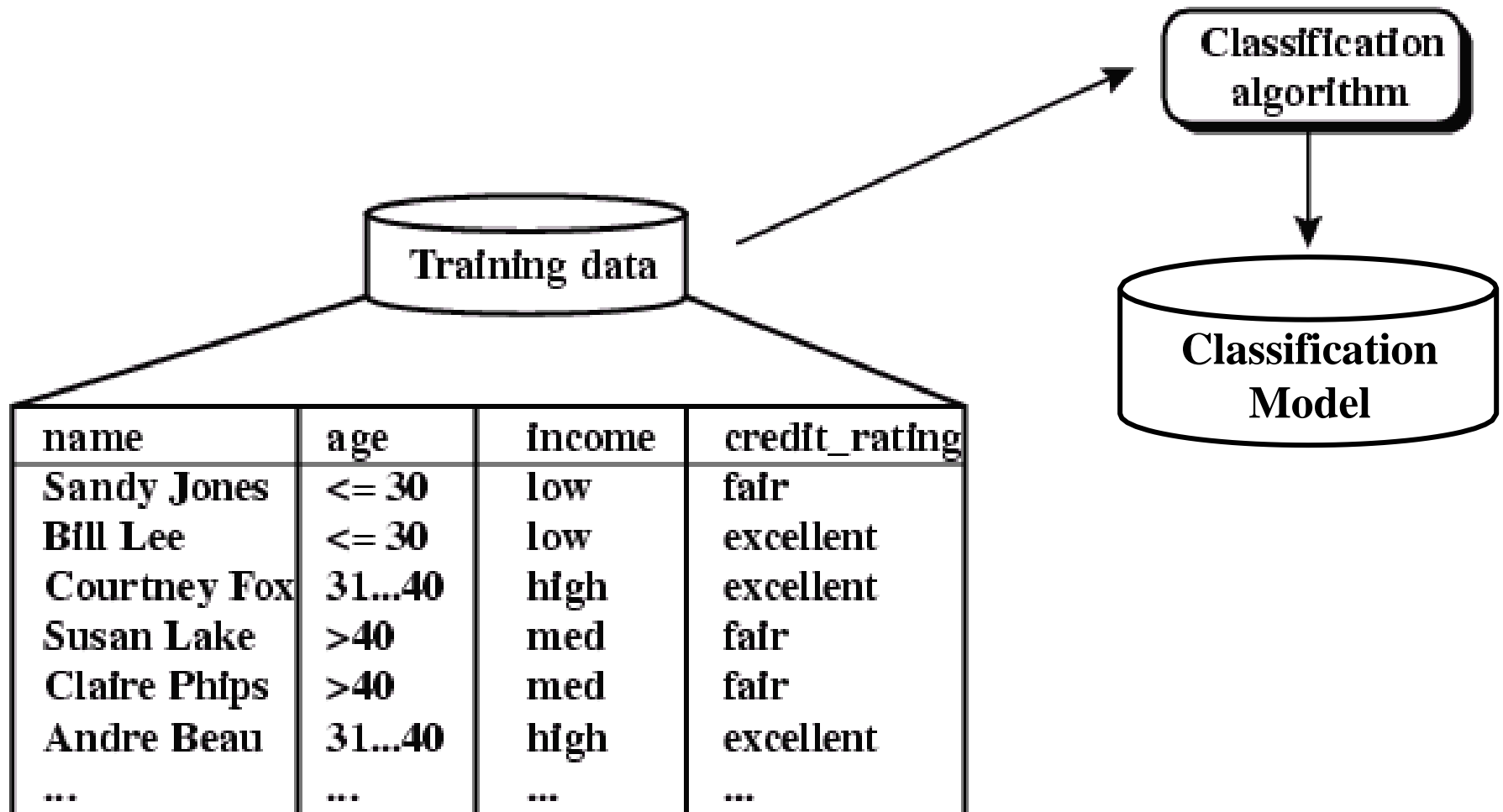
wucw@niu.edu.tw



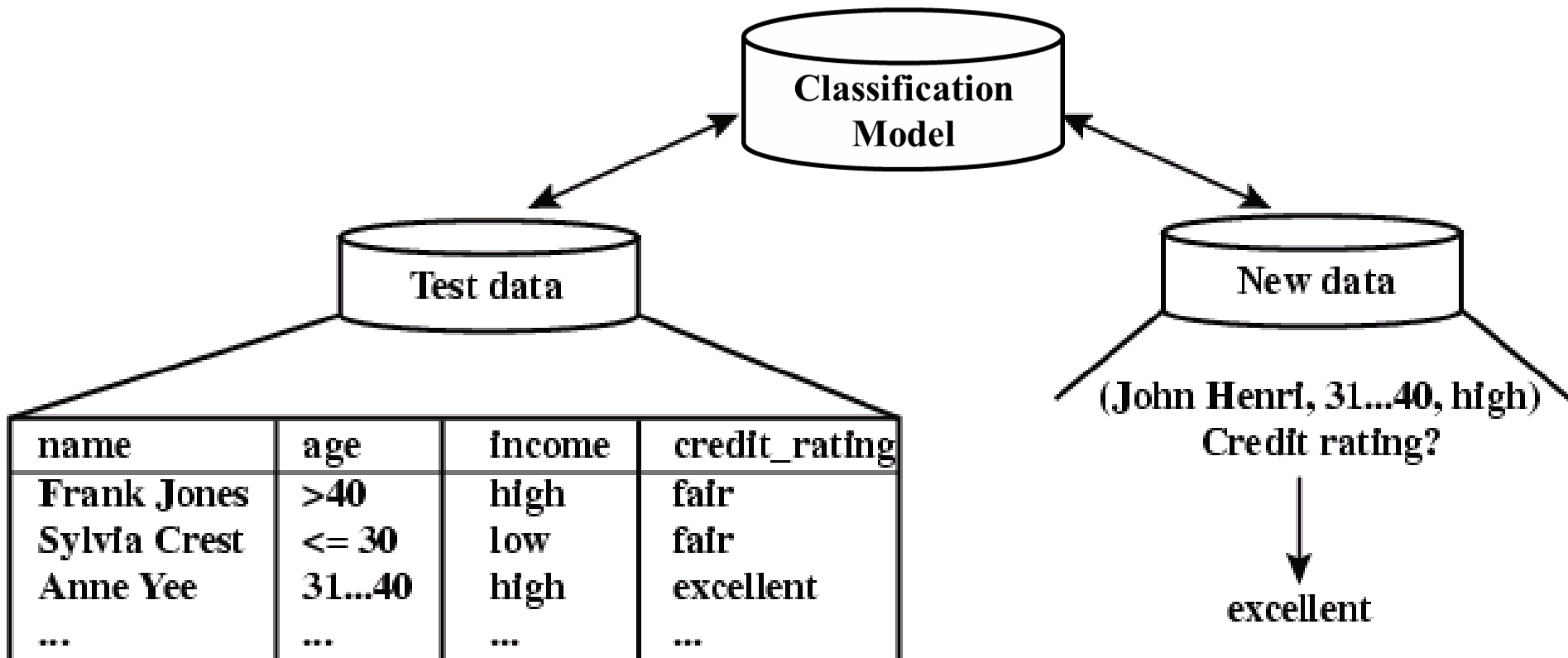
教學目標

- 了解二元分類模型的評估方法
- 了解多元分類模型的評估方法
- 了解訓練及測試資料分割方法及注意事項
- 了解不同評估方法的優缺點

Learning Phase



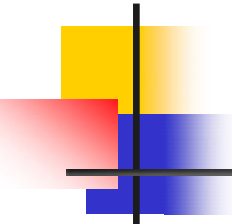
Testing Phase & Classification Phase





Evaluation Measurements for Binary Classifier

- 二元分類模型的評估方法
 - 正確率(Accuracy)
 - 精確率(Precision)
 - 捕捉率(Recall)
 - F-指標(F-measure)



比較測試資料之目標屬性與 模型分類結果

- 將測試資料輸入至分類模型後，可得到如下表的結果。

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

混亂矩陣(Confusion Matrix)

- 比較測試資料的真實目標屬性值與模型預測的結果
，可建構一個**混亂矩陣(Confusion Matrix)**。
- 混亂矩陣又稱為**列聯表(Contingency Table)**。

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

真實目標屬性值

	模型預測結果		Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

正確率計算方式

■ 正確率(Accuracy)

■ $(2+1)/8 = 0.375$

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

真實目標屬性值

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

準確率計算方式For Class {Yes}

■ 準確率(Precision) for Class {Yes}

■ $2/4 = 0.5$

準確率計算方式For Class {Yes}

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

真實目標屬性值

準確率計算方式For Class {No}

■ 準確率(Precision) for Class {No}

■ $1/4 = 0.25$

準確率計算方式For Class {No}

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

真實目標屬性值

捕捉率計算方式For Class {Yes}

■ 捕捉率(Recall) for Class {Yes}

■ $2/5 = 0.4$

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

真實目標屬性值

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

捕捉率計算方式For Class {No}

■ 捕捉率(Recall) for Class {No}

■ $1/3 = 0.33$

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

真實目標屬性值

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

F-指標計算方式

■ F-指標(F-measure)

- Precision 與 Recall 的調合平均數 (Harmonic Average)。

ID	Actual Class	Prediction
1	Yes	No
2	Yes	No
3	Yes	No
4	Yes	Yes
5	Yes	Yes
6	No	Yes
7	No	Yes
8	No	No

模型預測結果

	Yes	No	Total
Yes	2	3	5
No	2	1	3
Total	4	4	8

真實目標屬性值

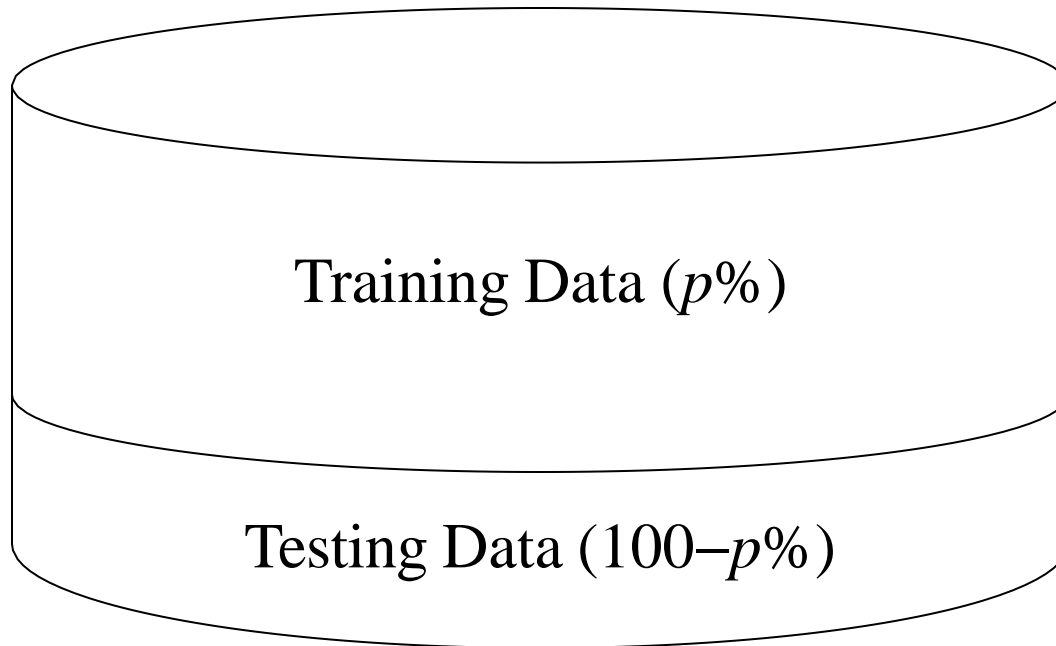


Separation of Training and Testing Data

- **Separation Methods**
 - 留出法(Holdout)
 - K疊交互驗證法(K-Fold Cross Validation)
 - 留一法(Leave-one-out)

留出法(Holdout)

- **Holdout method** samples $p\%$ of whole historical data as **training data**, and the rest $(100\% - p\%)$ data as **testing data**, where $1 \leq p \leq 100$.





留出法(Holdout)

- How to set p ?
 - 訓練資料太少，測試資料太多，可能導致模型不準確，拿再多測試資料測，還是不準確。
 - 訓練資料太多，測試資料太少，可能無法客觀評估所建模型的分類能力。
- In real-life applications, p is usually set to
 - $p = 80\%$ (80% training data, 20% testing data)
 - $p = 70\%$ (70% training data, 30% testing data)



留出法(Holdout)

- 留出法適用時機

- 當**資料量很大**時，資料集可能包含各種類型的資料。此時，以大量資料建立出來的模型，通常會具備不錯的準確率。此時可考慮使用留出法。

- 留出法主要優點

- 僅需建立一次模型的**資源成本(時間/硬碟/記憶體)**

- 留出法主要缺點

- 當資料量不大時，僅考慮一次抽樣建模的結果，**評估可能不夠客觀**。

K 疊交互驗證法

(K -Fold Cross Validation)

- Suppose that the classification method used is F .
- **K-fold cross validation method** separates the whole historical data D^* into K (nearly) equaled size of data $\{D_1, D_2, D_3, \dots, D_K\}$.
- For each dataset D_i ($1 \leq i \leq K$), taking dataset D_i as the **testing data** to build a classifier and the rest of data as the **training data** ($D^* - D_i$).



K 疊交互驗證法

(K -Fold Cross Validation)

- **K-fold cross validation method** leads to K classifiers.
- **In testing phase**, considering the overall classification performance of the K classifiers as the classification performance of the classification method F .
 - e.g., taking average accuracy of the K classifiers as the accuracy of F .
- **In on-line classification phase, using D^*** to build a final classifier for classification purpose.

K 疊交互驗證法

(K -Fold Cross Validation)

- **K -fold cross validation method** leads to K classifiers.
- **In testing phase**, considering the overall classification performance of the K classifiers as the classification performance of the classification method F .
 - e.g., taking average accuracy of the K classifiers as the accuracy of F .
- **In on-line classification phase, using D^*** to build a final classifier for classification purpose.



K 疊交互驗證法 (K -Fold Cross Validation)

- K 疊交互驗證法適用時機
 - 當**資料量中等**時，可使用 K 疊交互驗證法，建立 K 次分類模型，並以 K 次的整體分類結果，評估分類方法 F 。
- K 疊交互驗證法主要優點
 - 對分類方法 F 進行 K 次的模型評估，**評估結果較客觀**。
- K 疊交互驗證法主要缺點
 - 當**資料量很大**時，需**耗費** K 次模型建立的**資源成本**。



留一法(Leave-one-out)

- **Leave-one-out method** is a special case of K-fold cross validation method, where K is set to the size of D^* (i.e., $|D^*|$).



留一法(Leave-one-out)

■ 留一法適用時機

- 當**資料量很少**時，每一筆資料都相當珍貴。
- 例如在醫學領域，10個病人的資料得來不易，每一筆都相當珍貴。

■ 留一法主要優點

- 不會浪費任何一筆資料，每次建立模型都使用 $|D^*|-1$ 筆資料。

■ 留一法主要缺點

- 當**資料量多**時，需**耗費** $|D^*|$ 次模型建立的**資源成本**。

留出法搭配分層抽樣 (Holdout + Stratified Sampling)

