



巨匠線上真人

Python 資料科學應用開發

第十三堂：SVM 與決策樹

同學，歡迎你參加本課程

- ☑ 請關閉你的FB、Line等溝通工具，以免影響你上課。
- ☑ 考量頻寬、雜音，請預設關閉攝影機、麥克風，若有需要再打開。
- ☑ 隨時準備好，老師會呼叫你的名字進行互動，鼓勵用麥克風提問。
- ☑ 如果有緊急事情，你必需離開線上教室，請用聊天室私訊給老師，以免老師癡癡呼喚你的名字。
- ☑ 軟體安裝請在上課前安裝完成，未完成的同學，請盡快進行安裝。

課程檔案下載

The screenshot displays the homepage of the Juei Computer Online Live website. The header features navigation links for '開課查詢', '免費體驗專區', '課程總覽', '專業師資', '學員專區', '講師專區', and '最新消息'. Social media icons for Line, Facebook, and YouTube are also present. A user greeting '您好!' and a '登出' button are visible. The main banner area contains a large text overlay: '程式語言好難學? 那是因為你還沒學過Python! (線上老師 LIVE 直播教學 · 搶先看)'. On the right, a dropdown menu is open, listing various user options. The '課程檔案下載' option is highlighted with an orange box and a callout bubble. The background of the page features a blue and orange abstract design with circuit-like patterns.

巨匠電腦線上真人 開課查詢 免費體驗專區 課程總覽 專業師資 學員專區 講師專區 最新消息

360 f YouTube

您好! 登出

點數卡產品兌換
APCS檢測專區
公告專區
我的課表
IT真人課程劃位
電腦分校課程劃位
外語真人課程劃位
美語分校課程劃位
取消劃位
課程檔案下載
上課權益查詢
教學平台測試
學習諮詢
常見問題
個資維護
忘記密碼
登出

課程檔案下載

程式語言好難學?
那是因為
你還沒學過Python!
(線上老師 LIVE 直播教學 · 搶先看)

巨匠電腦真人課程

ZOOM 學員操作說明

The screenshot shows the Zoom interface with several callouts:

- 5 查看選項/共同註記/筆 (連連看)**: Points to the '共同註記' (Co-Annotate) option in the '查看選項' (View Options) menu.
- 2 共享螢幕 (指導演練；點評作品)**: Points to the '共享螢幕' (Share Screen) button in the bottom toolbar. A sub-note says: '老師須先停止共享螢幕才能請學生共享螢幕' (The teacher must first stop sharing the screen before asking the student to share the screen).
- 1 聊天**: Points to the '聊天' (Chat) button in the bottom toolbar.
- 3 與會者/舉手**: Points to the '與會者' (Participants) button in the bottom toolbar. A sub-note says: '舉手' (Raise Hand).
- 4 解除靜音**: Points to the '解除靜音' (Unmute) button in the bottom toolbar.

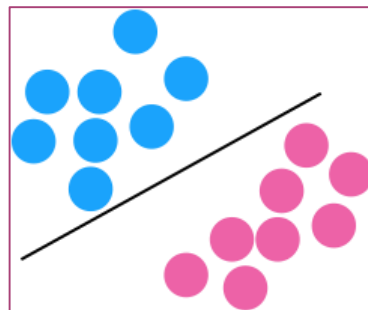
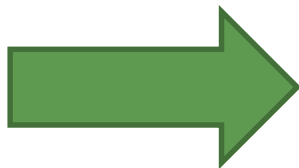
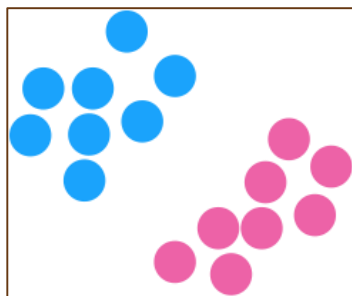
The interface also shows a list of participants in the top left, including '張齡月 (我)' and '婷婷'.

分類 (Classification)

- ◆ 最近距離分群法 (KNN)
- ◆ 羅吉斯迴歸 (Logistic Regression)
- ◆ **支持向量機 (Support Vector Machine)**
- ◆ 決策樹 (Decision Tree)
- ◆ 隨機森林 (Random Forest)

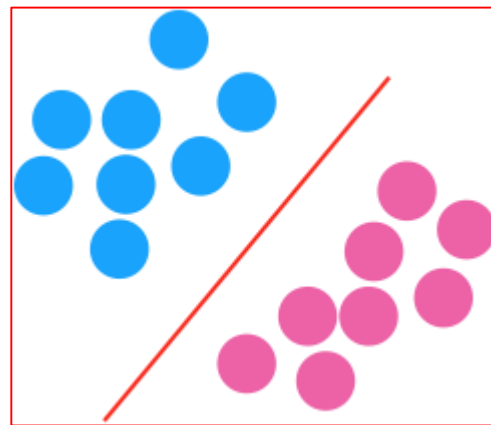
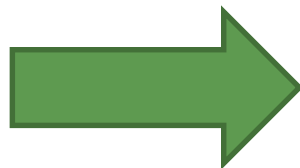
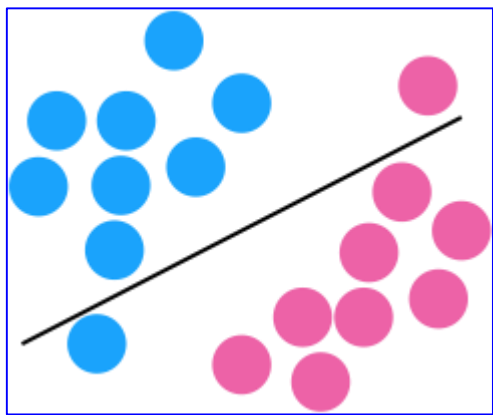
SVM 說明

- ◆ 我們可依照 Perceptron 或 Logistic Regression 將兩種不同顏色的球進行分類。



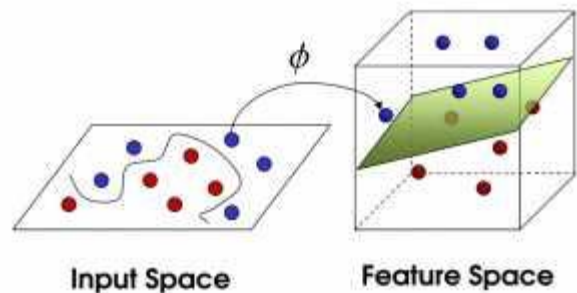
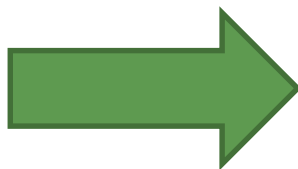
SVM 說明

- ◆ 可是你會發現當球增加後原本規劃的分類會產生問題，所以開始進行微調。



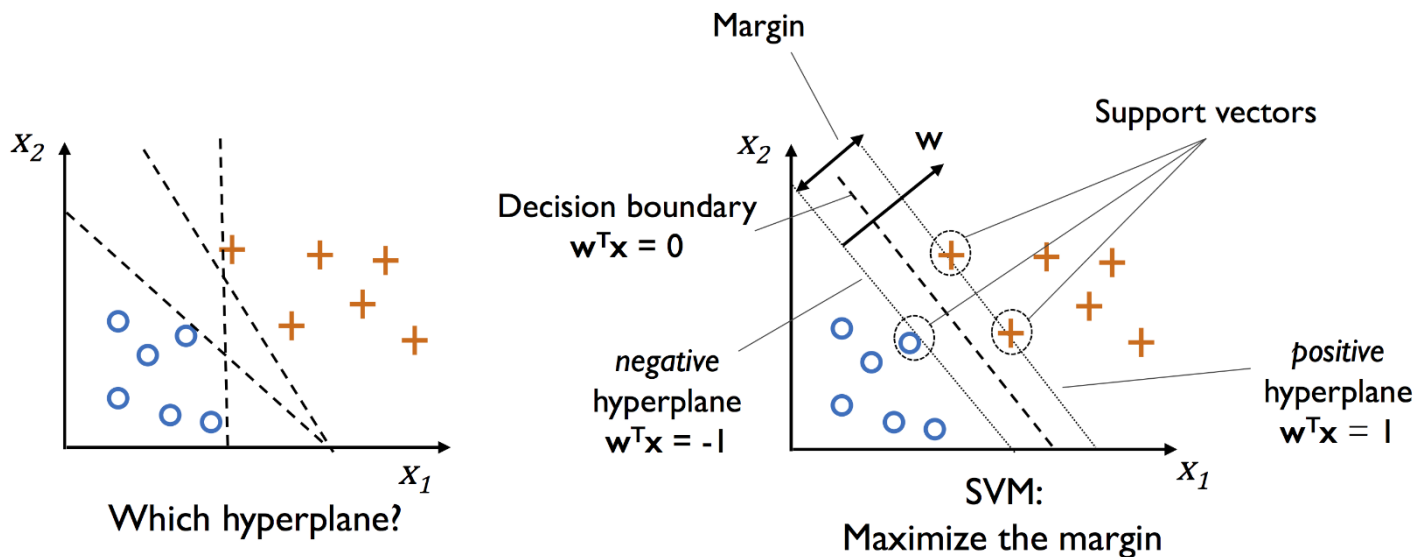
SVM 說明

- ◆ 若資料已經不容易分類，我們是否可以改用非線性的分類方式呢？



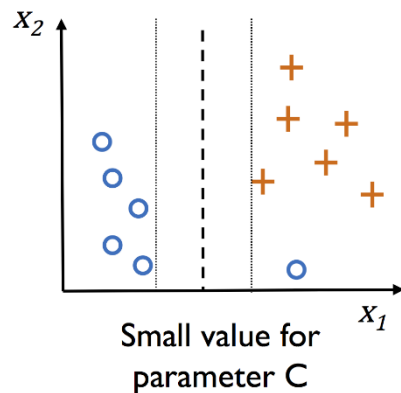
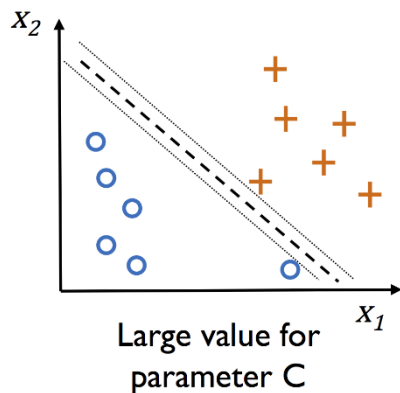
支援向量機 (Support Vector Machine ; SVM)

- ◆ 找到一完美的超平面 (Hyper plane) ，使切割的 margin 愈寬愈好



邊界 (Margin) 寬度的調整

- ◆ SVM 的參數 C ，可以控制錯誤分類的懲罰 (Penalty)
 - ◇ C 愈大，表對錯誤分類的懲罰愈大，即擬合度愈佳



實作

- ◆ 程式碼

- ◆ svm.py

- ◆ 測試

- ◆ 調整 C

SVM 優缺點

◆ SVM 優點

- ◆ 切出來的線很漂亮，擁有最大 margin 的特性。
- ◆ 可以很容易透過更換 Kernel，做出非線性的線（非線性的決策邊界）。

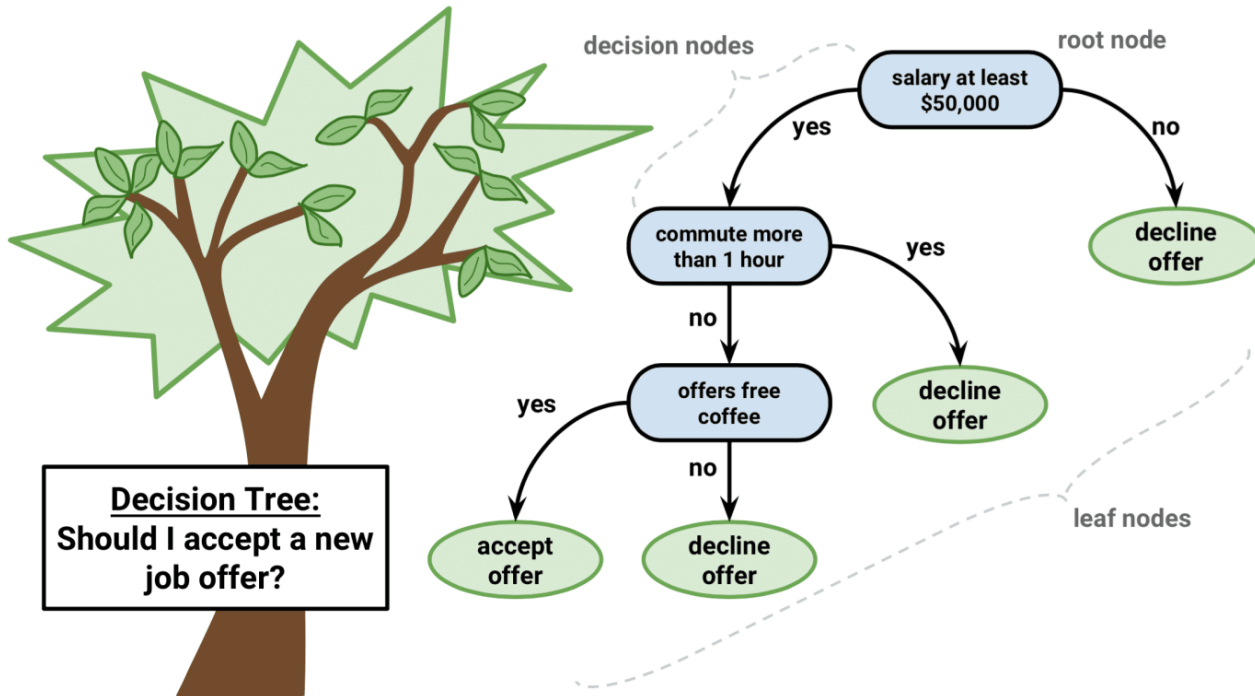
◆ SVM 缺點

- ◆ 效能較不佳，由於時間複雜度為 $O(n^2)$ 當有超過一萬筆資料時，運算速度會慢上許多。

分類 (Classification)

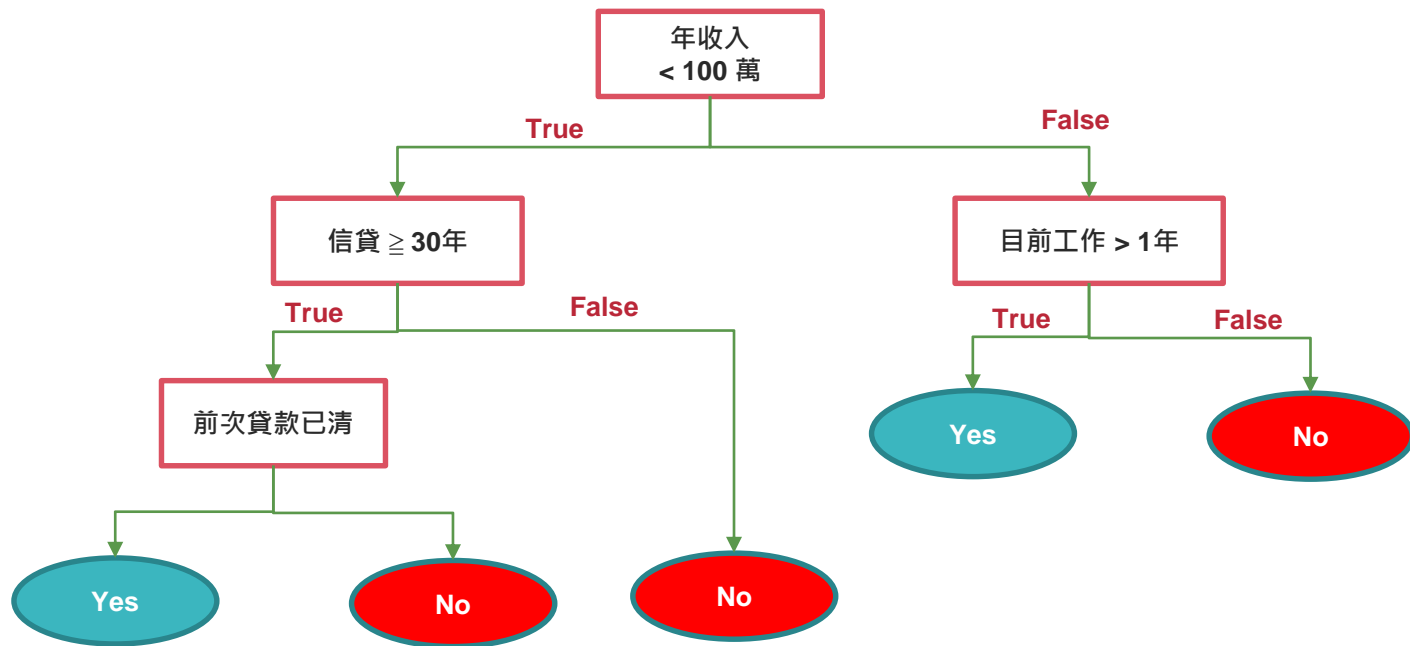
- ◆ 最近距離分群法 (KNN)
- ◆ 羅吉斯迴歸 (Logistic Regression)
- ◆ 支持向量機 (Support Vector Machine)
- ◆ **決策樹 (Decision Tree)**
- ◆ 隨機森林 (Random Forest)

決策樹 (Decision Tree)



決策樹 (Decision Tree)

◆ 依特徵值的重要性逐步分類



特徵值的重要性

- ◆ 分類基準：資訊增益 (Information Gain) ，以資訊量衡量特徵的重要性，常見的資訊量：
 - ◆ 熵 (Entropy)
 - ◆ Gini 不純度 (Gini Impurity)
- ◆ 資訊增益 (Information Gain) : The expected information gain is the change in entropy
 - ◆ $Entropy = -\sum_i -p_i \log_2 p_i$
 - ◆ $Information\ Gain = entropy(parent) - [average\ entropy(children)]$
 - $IG(T, a) = H(T) - H(T|a)$

演算法進行步驟

- ◆ 資料設定：將原始資料分成兩組，一部分為訓練資料，一部分為測試資料
- ◆ 決策樹生成：使用訓練資料來建立決策樹，而在每一個內部節點，則依據屬性選擇指標〔如：資訊理論 (Information Theory) ...〕來評估選擇哪個屬性做分支的依據。此又稱節點分割 (Splitting Node)
- ◆ 剪枝 (Tree pruning)：使用測試資料來進行決策樹修剪，將以上1 ~ 3步驟不斷重複進行，直到所有的新產生節點都是樹葉節點為止
- ◆ 防止 Over fitting (過度擬合) 的方法：
 - ◆ Minimum samples for a node split：資料數目不得小於多少才能再產生新節點
 - ◆ Minimum samples for a terminal node (leaf)：要成為葉節點，最少需要多少資料
 - ◆ Maximum depth of tree (vertical depth)：限制樹的高度最多幾層
 - ◆ Maximum number of terminal nodes：限制最終葉節點的數目
 - ◆ Maximum features to consider for split：在分離節點時，最多考慮幾種特徵值

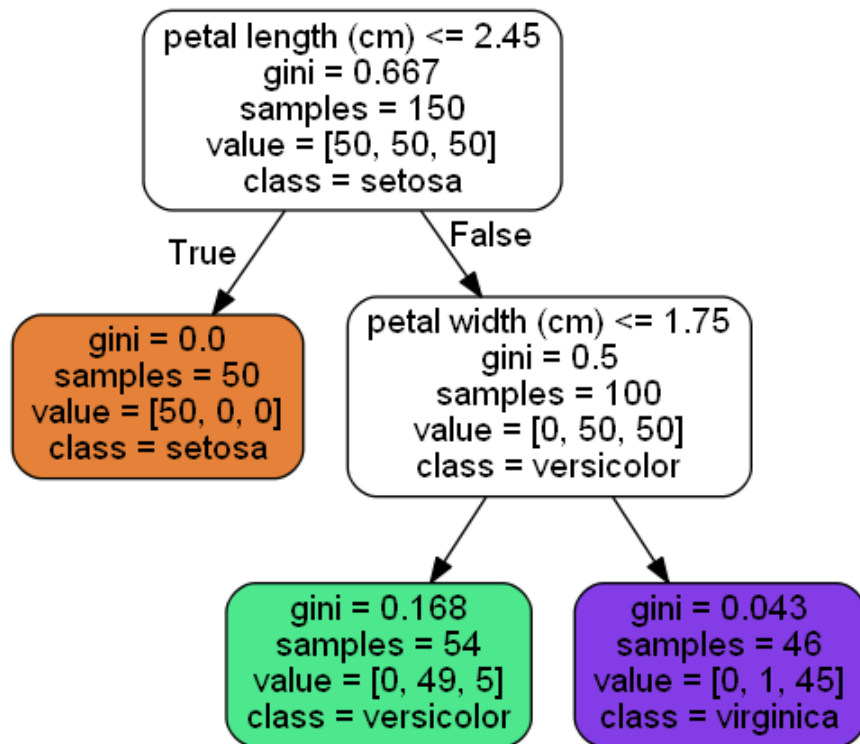
實作

◆ 程式碼

◆ decisionTree_test.py

◆ 測試

◆ 調整 max_depth

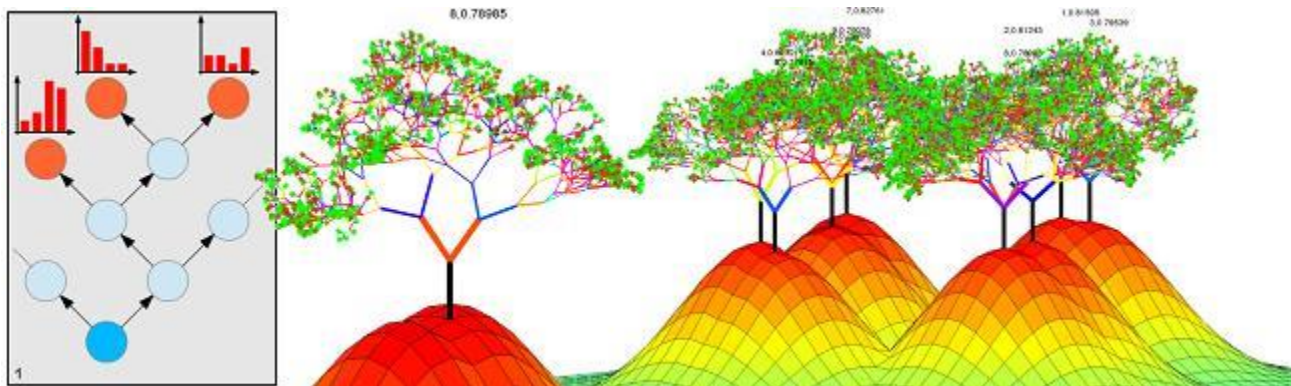


分類 (Classification)

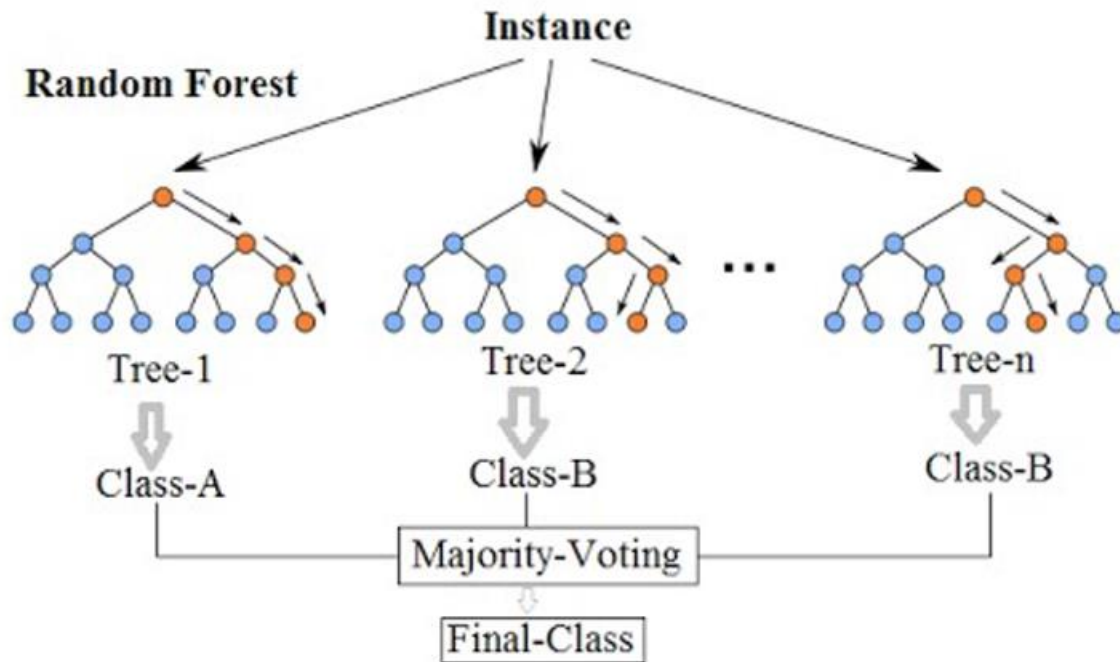
- ◆ 最近距離分群法 (KNN)
- ◆ 羅吉斯迴歸 (Logistic Regression)
- ◆ 支持向量機 (Support Vector Machine)
- ◆ 決策樹 (Decision Tree)
- ◆ **隨機森林 (Random Forest)**

隨機森林 (Random Forest)

- ◆ 用隨機的方式建立很多個的決策樹，每一棵樹都會給出自己的分類選擇，並由此進行「投票」，森林整體的輸出結果將會是票數最多的類別



多數決



<https://www.youtube.com/watch?v=ajTc5y3OqSQ>

實作

- ◆ 程式碼

 - ◆ RandomForest.py

- ◆ 測試

 - ◆ 調整 `n_estimators` (`n` 棵樹) 、 `max_depth` (`m` 層)

作品：以決策樹分析鐵達尼資料集

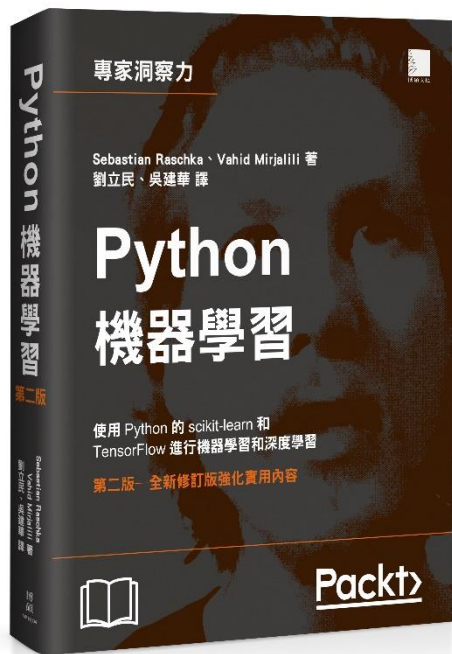
- ◆ 資料集：鐵達尼 (Titanic)
- ◆ `import seaborn as sns`
- ◆ `titanic = sns.load_dataset("titanic")`
- ◆ 參考

展示



<https://blogs.msdn.microsoft.com/mengtsai/2015/03/11/943/>

參考用書



- ◆ 書名：Python機器學習（第二版）

<http://www.drmaster.com.tw/bookinfo.asp?BookID=MP11804>

- ◆ 作者：Sebastian Raschka, Vahid Mirjalili ISBN
- ◆ 譯者：劉立民、吳建華
- ◆ 出版社：博碩

問卷

<http://www.pcschoolonline.com.tw>

開課查詢

免費體驗專區

課程總覽

專業師

1

學員專區

講師專區



➤ 課程檔案下載：

學員的「上課教材」，下載檔案為壓縮檔 ([解壓縮操作步驟](#))。
如無法觀看上課教材，請安裝 [PDF閱讀軟體](#)。

公告專區

我的課表

課程劃位

取消劃位

2

課程檔案下載

自107年1月1日起，課程錄影檔由180天改為365天(含)內無限次觀看 (上課隔日18:00起)。

問
卷

上課日期	課程名稱	課程節次	教材下載		
2017/12/27 2000 ~ 2200	線上真人-ZBrush 3D動畫造型設計	18	上課教材	錄影檔	課堂問卷
2017/12/20 2000 ~ 2200	線上真人-ZBrush 3D動畫造型設計	17	上課教材	錄影檔	
2017/12/18 2000 ~ 2200	線上真人-ZBrush 3D動畫造型設計	16	上課教材	錄影檔	



巨匠線上真人

www.pcschoolonline.com.tw