FB: <https://www.facebook.com/groups/2032749180362722/>

Cupoy: <https://ai100-2.cupoy.com/>

## D31：機器學習概論

### Sample Code & 作業內容

請觀看台大電機李宏毅教授的影片ML Lecture 0-1: Introduction of Machine Learning。李教授會用非常淺顯易懂的方式介紹甚麼是機器學習。看完影片請回答問題。

作業問題：請找出一個你認為機器學習可以發光發熱的領域並闡述原因?

請點擊檢視範例，作業請提交Day\_031\_HW.ipynb

影片連結：[**https://www.youtube.com/watch?v=CXgbekl66jc**](https://www.youtube.com/watch?v=CXgbekl66jc)

### 參考資料

### 了解機器學習/人工智慧的各種名詞意義

連結：[**機器是如何學習到東西的?**](https://kopu.chat/2017/07/28/%E6%A9%9F%E5%99%A8%E6%98%AF%E6%80%8E%E9%BA%BC%E5%BE%9E%E8%B3%87%E6%96%99%E4%B8%AD%E3%80%8C%E5%AD%B8%E3%80%8D%E5%88%B0%E6%9D%B1%E8%A5%BF%E7%9A%84%E5%91%A2/)

### 聽聽人工智慧頂尖學者分享 AI 的知識

連結：[**Stanford 李飛飛教授 TED Talk - 如何教懂電腦看圖像**](https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?language=zh-tw)

## D32：機器學習-流程與步驟

### Sample Code & 作業內容

請至以下機器學習巨頭的 blog 參觀，試著閱讀一篇文章並分析其

作業 1：目標

作業 2：方法

作業 3：資料來源

* [**Google AI blog**](https://ai.googleblog.com/)
* [**Facebook Research blog**](https://research.fb.com/blog/)
* [**Apple machine learning journal**](https://machinelearning.apple.com/)
* [**機器之心**](https://www.jiqizhixin.com/)
* [**雷鋒網**](https://www.leiphone.com/category/ai)

請點擊檢視範例，作業請提交Day\_032\_HW.ipynb

### 參考資料

ML 流程 by Google-影片連結：[**The 7 Steps of Machine Learning (AI Adventures)**](https://www.youtube.com/watch?v=nKW8Ndu7Mjw)

## D33：機器如何學習?

### Sample Code & 作業內容

請觀看李宏毅教授以神奇寶貝進化 CP 值預測的範例，解說何謂機器學習與過擬合。

影片連結：[**ML Lecture 1: Regression - Case Study**](https://www.youtube.com/watch?v=fegAeph9UaA)

看完影片請回答以下問題：

1. 模型的泛化能力 (generalization) 是指什麼？

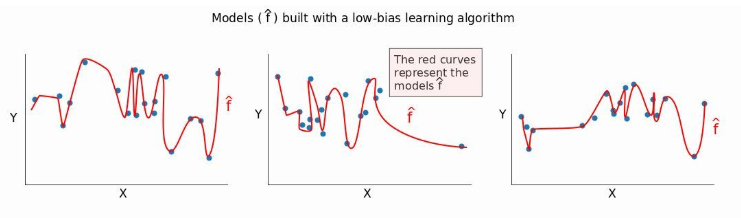
2. 分類問題與回歸問題分別可用的目標函數有哪些？

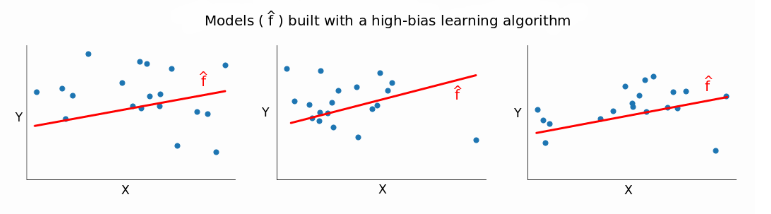
請點擊檢視範例，作業請提交Day\_033\_HW.ipynb

### 參考資料

### 理解機器學習中很重要的 Bias/Variance trade-off 的意義為何

影片連結：[**學習曲線與 bias/variance trade-off**](http://bangqu.com/yjB839.html)





## D34：訓練/測試集切分的概念

### Sample Code & 作業內容

作業：類別平衡切分

現實中的分類問題經常遇到類別不均衡的問題。如瑕疵分類，瑕疵樣本通常總是比正常樣本少。

請參考範例程式碼Day\_034\_train\_test\_split.ipynb試著撰寫程式碼，讓測試集的目標類別能夠均衡分佈。

作業請提交Day\_034\_HW.ipynb

### 參考資料

### 理解訓練、驗證與測試集的意義與用途

影片連結：[**台大電機李宏毅教授講解訊練/驗證/測試集的意義**](https://www.youtube.com/watch?v=D_S6y0Jm6dQ&feature=youtu.be&t=1948)

## D35：regression vs. classification

### Sample Code & 作業內容

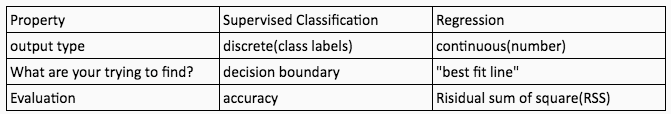
請搜尋目標為多標籤問題 (Multi-label) 的機器學習案例 (Hint: 服飾)，了解其資料來源、目標以及評估指標為何

請點擊檢視範例，作業請提交Day\_035\_HW.ipynb

### 參考資料

### 了解回歸與分類的差異在哪裡?

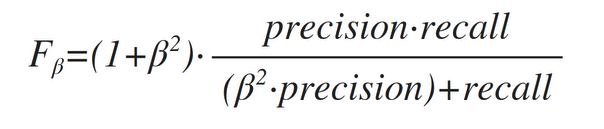
網頁連結：[**回歸與分類的比較**](http://zylix666.blogspot.com/2016/06/supervised-classificationregression.html)



## D36：評估指標選定/evaluation metrics

### Sample Code & 作業內容

* F1-Score 其實是 F-Score 中的 β 值為 1 的特例，代表 Precision 與 Recall 的權重相同
* 實際上可以使用 F2-Score，代表 β 值為 2，公式如下，請試著撰寫 F2-Score 的程式碼



    作業請提交Day\_036\_HW.ipynb

### 參考資料

* 深入了解超常用的指標 AUC

         網頁連結：[**超詳細解說 AUC (英文)**](https://www.dataschool.io/roc-curves-and-auc-explained/)

* 學習更多評估指標，來衡量機器學習模型的準確度

         網頁連結：[**更多評估指標**](https://zhuanlan.zhihu.com/p/30721429)

## D37：regression model 介紹 - 線性迴歸/羅吉斯回歸

### Sample Code & 作業內容

請閱讀以下相關文獻並試著回答以下問題 (本日知識點建議閱讀補充資料)

* [**Linear Regression 詳細介紹**](https://brohrer.mcknote.com/zh-Hant/how_machine_learning_works/how_linear_regression_works.html)
* [**Logistics Regression 詳細介紹**](https://medium.com/@yehjames/%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%AC3-3%E8%AC%9B-%E7%B7%9A%E6%80%A7%E5%88%86%E9%A1%9E-%E9%82%8F%E8%BC%AF%E6%96%AF%E5%9B%9E%E6%AD%B8-logistic-regression-%E4%BB%8B%E7%B4%B9-a1a5f47017e5)
* [**你可能不知道的 Logistic Regression**](https://taweihuang.hpd.io/2017/12/22/logreg101/)

線性回歸模型能夠準確預測非線性關係的資料集嗎?

回歸模型是否對資料分布有基本假設?

作業請提交 Day\_037\_HW.ipynb

### 參考資料

* [**超人氣 Stanford 教授 Andrew Ng 教你 Linear regression**](https://zh-tw.coursera.org/lecture/machine-learning/model-representation-db3jS) (強烈推薦觀看)
* [**Logistic regression 數學原理**](https://blog.csdn.net/qq_23269761/article/details/81778585)

## D38：regression model 程式碼撰寫

### Sample Code & 作業內容

作業１：請改用 sklearn.datasets 中除了範例以外至少兩組 datasets 來訓練自己的回歸模型 (根據不同資料的選擇適合的模型)

作業２：練習完建議參考補充資料，觀看更多回歸模型的範例

請參考範例程式碼 Day\_038\_regression\_model.ipynb，作業請提交Day\_038\_HW.ipynb

### 參考資料

* [**超多 Linear Regression / Logistic Regression 的 examples**](https://github.com/trekhleb/homemade-machine-learning)
* [**深入了解 multinomial Logistic Regression 的原理**](http://dataaspirant.com/2017/05/15/implement-multinomial-logistic-regression-python/)

## D39：regression model 介紹 - LASSO 回歸/ Ridge 回歸

### Sample Code & 作業內容

請閱讀相關文獻，並回答下列問題：

* [**脊回歸 (Ridge Regression)**](https://blog.csdn.net/daunxx/article/details/51578787)
* [**Linear, Ridge, Lasso Regression 本質區別**](https://www.zhihu.com/question/38121173)

作業１：LASSO 回歸可以被用來作為 Feature selection 的工具，請了解 LASSO 模型為什麼可用來作 Feature selection

作業２：當自變數 (X) 存在高度共線性時，Ridge Regression 可以處理這樣的問題嗎?

作業請提交Day\_039\_HW.ipynb

## D40：regression model 程式碼撰寫

### Sample Code & 作業內容

作業：嘗試至少兩個不同的資料集，並調整 α 值來觀察模型訓練的情形

請參考範例程式碼Day\_040\_lasso\_ridge\_regression.ipynb，作業請提交Day\_040\_HW.ipynb

## D41：tree based model - 決策樹 (Decision Tree) 模型介紹

### Sample Code & 作業內容

閱讀以下兩篇文獻，了解決策樹原理，並試著回答後續的問題

* [**決策樹 (Decision Tree) - 中文**](https://medium.com/@yehjames/%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%AC3-5%E8%AC%9B-%E6%B1%BA%E7%AD%96%E6%A8%B9-decision-tree-%E4%BB%A5%E5%8F%8A%E9%9A%A8%E6%A9%9F%E6%A3%AE%E6%9E%97-random-forest-%E4%BB%8B%E7%B4%B9-7079b0ddfbda)
* [**how decision tree works - 英文**](http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/)

作業１：在分類問題中，若沒有任何限制，決策樹有辦法在訓練時將 training loss 完全降成 0 嗎？

作業２：決策樹做分類問題時，資料的相似度比較容易計算 (是否屬於同一個類別)。那如果變成回歸問題，這時切分後的資料不純度該如何計算？樹建置完成後，又該如何進行預測呢？

作業請繳交Day\_041\_HW.ipynb

## D42：tree based model - 決策樹程式碼撰寫

### Sample Code & 作業內容

請參考範例程式碼Day\_042\_decision\_tree.ipynb

作業1：試著調整 DecisionTreeClassifier(...) 中的參數，並觀察是否會改變結果？

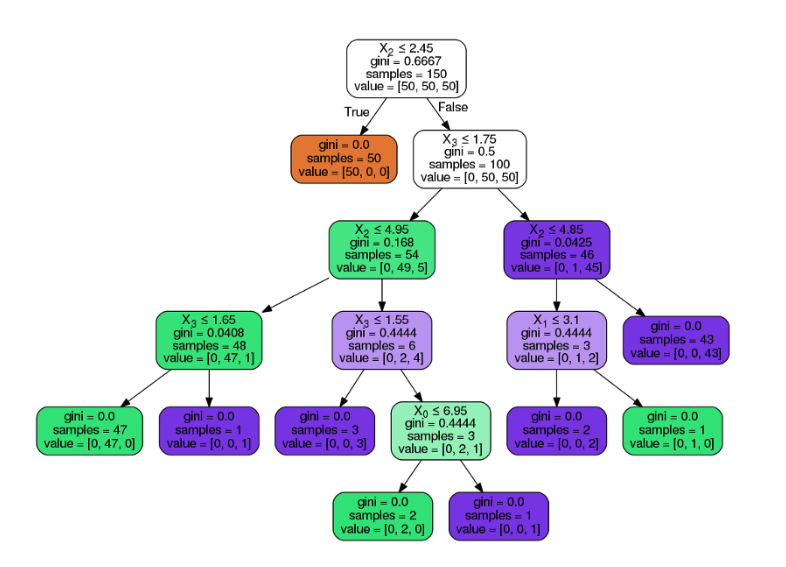
作業2：改用其他資料集 (boston, wine)，並與回歸模型的結果進行比較

作業請提交Day\_042\_HW.ipynb

### 參考資料

#### 可安裝額外的套件 graphviz，畫出決策樹的圖形幫助理解模型分類的準則

[**網頁連結**](https://medium.com/@rnbrown/creating-and-visualizing-decision-trees-with-python-f8e8fa394176)



## D43：tree based model - 隨機森林 (Random Forest) 介紹

### Sample Code & 作業內容

閱讀以下兩篇文獻，了解隨機森林原理，並試著回答後續的思考問題

* [**隨機森林 (random forest) - 中文**](http://hhtucode.blogspot.com/2013/06/ml-random-forest.html)
* [**how random forest works - 英文**](https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674)

1. 隨機森林中的每一棵樹，是希望能夠

* 沒有任何限制，讓樹可以持續生長 (讓樹生成很深，讓模型變得複雜)
* 不要過度生長，避免 Overfitting

2. 假設總共有 N 筆資料，每棵樹用取後放回的方式抽了總共 N 筆資料生成，請問這棵樹大約使用了多少 % 不重複的原資料生成?

* hint: 0.632 bootstrap

作業請提交Day\_043\_HW.ipynb

## D44：tree based model - 隨機森林程式碼撰寫

### Sample Code & 作業內容

請參考範例程式碼Day\_044\_random\_forest.ipynb

作業1：試著調整 RandomForestClassifier(...) 中的參數，並觀察是否會改變結果？

作業2：改用其他資料集 (boston, wine)，並與回歸模型與決策樹的結果進行比較

作業請提交Day\_044\_HW.ipynb

### 參考資料

知名 ML youtuber 教你手刻隨機森林 by Python

影片連結：[**https://www.youtube.com/watch?v=QHOazyP-YlM**](https://www.youtube.com/watch?v=QHOazyP-YlM)

## D45：tree based model - 梯度提升機 (Gradient Boosting Machine) 介紹

### Sample Code & 作業內容

你可能聽過 XGBoost/Light-GBM，這些都是資料科學競賽中最常用的機器學習模型，但其實這些演算法背後原理都是基於 Gradient-boosting 進而優化，強烈建議您對本日的課程與補充教材多花點時間閱讀與理解。 核心概念就是透過計算梯度，來讓下一棵生成的樹能夠根據梯度方向，試圖讓 Loss 變得更小！

 本日作業請完整閱讀以下任一文獻即可：

* [**Kaggle 大師帶你了解梯度提升機原理 - 英文**](http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/)
* [**完整的 Ensemble 概念 by 李宏毅教授**](https://www.youtube.com/watch?v=tH9FH1DH5n0)
* [**深入了解 Gradient-boosting - 英文**](https://explained.ai/gradient-boosting/index.html)

完成閱讀後，請記得到下方按下完成作業。

### 參考資料

#### 梯度提升機原理 - 中文

連結：[**https://ifun01.com/84A3FW7.html**](https://ifun01.com/84A3FW7.html)

文章中的殘差就是前面提到的 Loss，從範例中了解殘差是如何被修正的

#### XGboost 作者講解原理 - 英文

連結：[**https://www.youtube.com/watch?v=ufHo8vbk6g4**](https://www.youtube.com/watch?v=ufHo8vbk6g4)

了解 XGBoost 的目標函數是什麼，模型是怎麼樣進行優化

#### XGBoost 數學原理 slides - 英文

連結：[**https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf**](https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf)

了解 XGBoost 的目標函數數學推導

## D46：tree based model - 梯度提升機程式碼撰寫

### Sample Code & 作業內容

目前已經學過許多的模型，相信大家對整體流程應該比較掌握了，這次作業請改用手寫辨識資料集，步驟流程都是一樣的，請試著自己撰寫程式碼來完成所有步驟。請參考範例程式碼Day\_046\_gradient\_boosting\_machine.ipynb，作業請提交Day\_046\_HW.ipynb。

### 參考資料

#### 完整調參數攻略-如何使用 Python 調整梯度提升機的超參數

[**complete-guide-parameter-tuning-gradient-boosting-gbm**](https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/)