

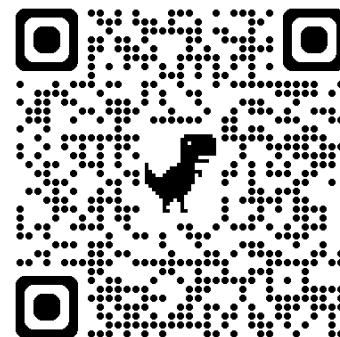


# 深度學習

# Deep Learning

## Project Introduction

Instructor: 林英嘉 (Ying-Jia Lin)  
2025/04/07

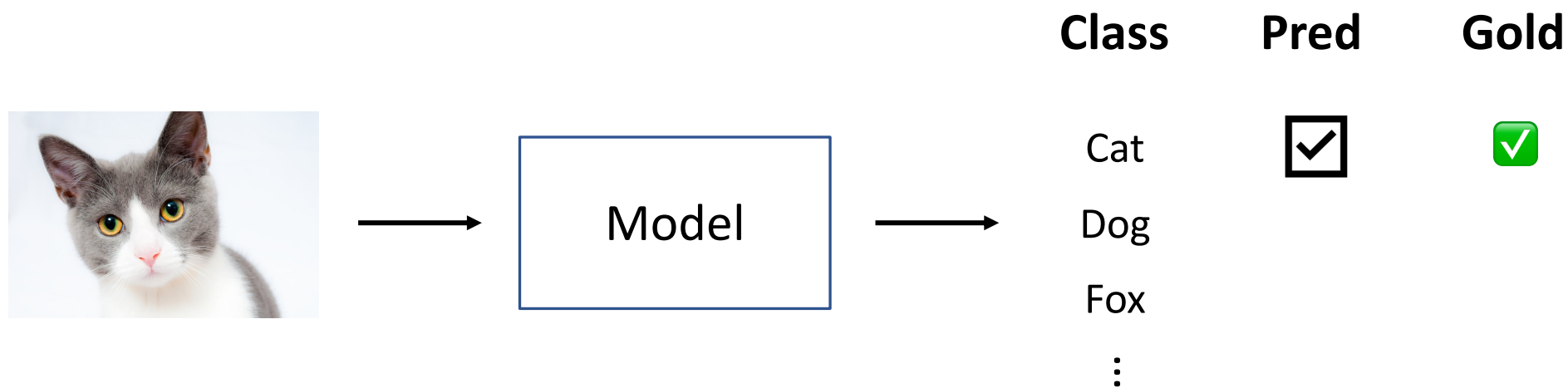


[Course GitHub](#)

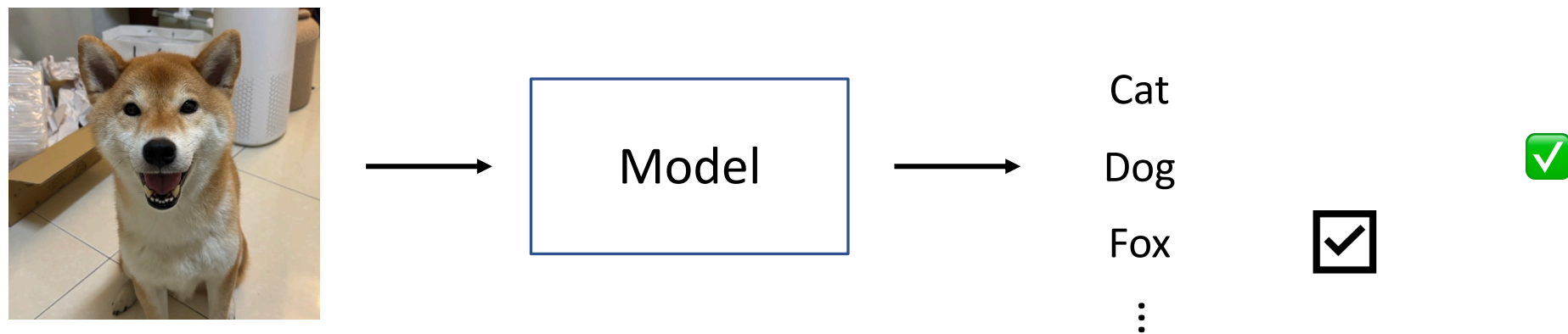


[Slido # DL\\_0324](#)

# 如何評估模型效能？

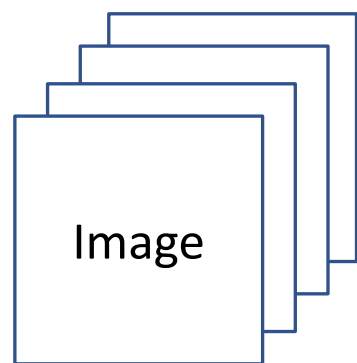


我們不能只憑一筆資料就說模型很棒...



# 自動化評估

- 「評估」的意義：量化模型的表現



資料



(假設用二分類任務)

	Pred	Gold
1	1	0
2	1	1
3	1	1
4	1	1

正確率：0.75

真正答對的數量 /  
總預測的數量



# 評估方式 (計算模型有多高分的方式)

---

- 不同任務有不同的評估方式

類型

Automatic Evaluation  
(自動化評估)



Human Evaluation  
(人工評估)



「生成」任務較為需要



# What is Kaggle?

---

- Kaggle is a platform that provides:
  - Real-world datasets for machine learning
  - Competitions with prizes (sometimes with money)
  - Discussion forum with **a lots of code examples**

kaggle



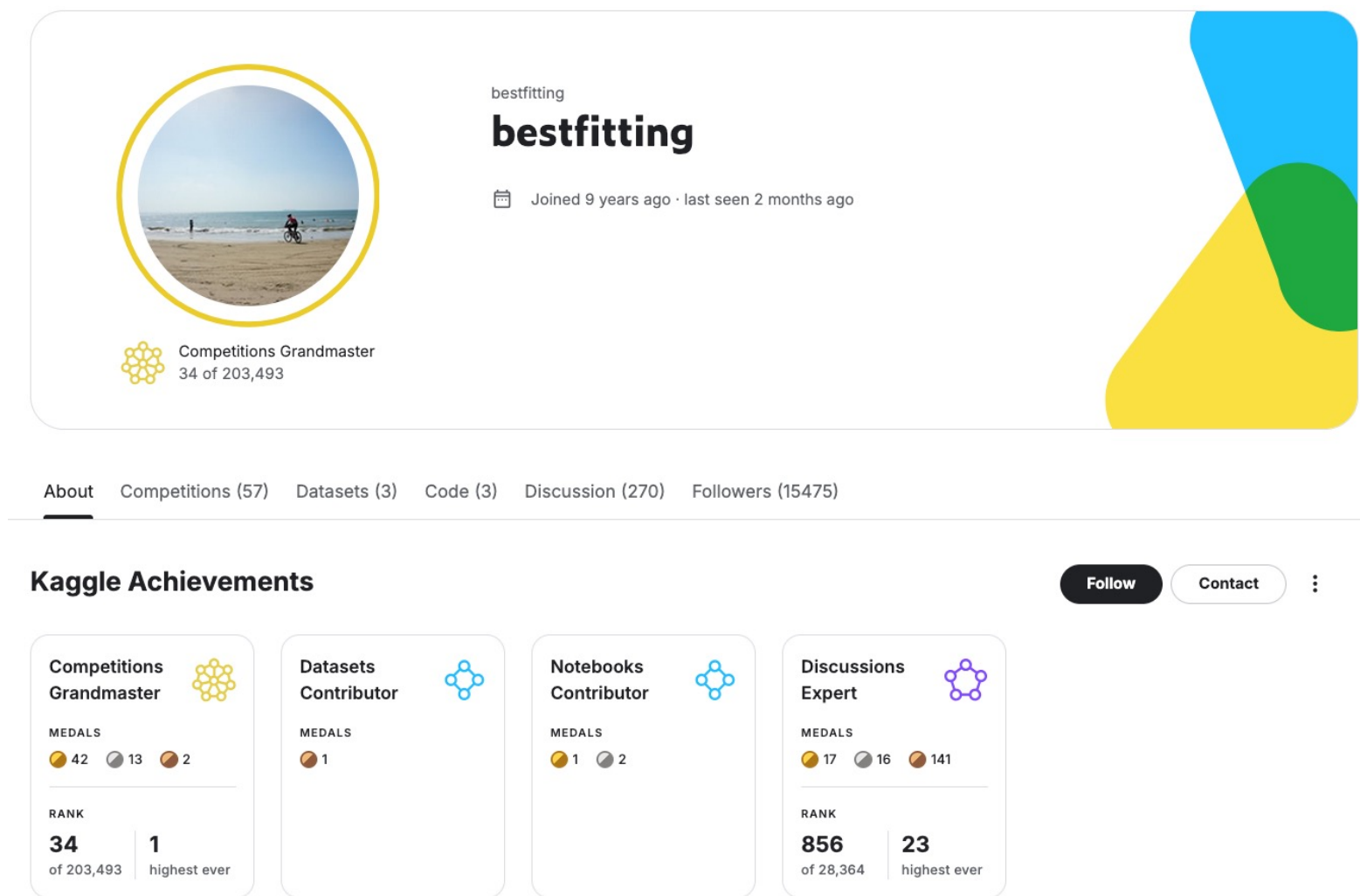
# Kaggle submission types

---

- Traditional competitions 本課程 projects 採用此方式
  - Upload **submission file** (e.g., \*.csv)
- Code / Notebook competitions
  - Upload **code** (e.g., \*.ipynb)



# If you play Kaggle a lot ...



The image shows a Kaggle user profile for 'bestfitting'. The profile includes a circular profile picture of a beach scene, a yellow border, and a 'Competitions Grandmaster' badge (34 of 203,493). The user's name 'bestfitting' is displayed, along with their join date (9 years ago) and last seen date (2 months ago). A navigation bar shows links to 'About', 'Competitions (57)', 'Datasets (3)', 'Code (3)', 'Discussion (270)', and 'Followers (15475)'. Below this is a 'Kaggle Achievements' section with four cards: 'Competitions Grandmaster' (Rank 34 of 203,493, 1 highest ever), 'Datasets Contributor' (Rank 1 of 28,364, 23 highest ever), 'Notebooks Contributor' (Rank 856 of 28,364, 23 highest ever), and 'Discussions Expert' (Rank 23 of 28,364, 23 highest ever). Each card shows medal counts (Gold, Silver, Bronze) and a 'Follow' button.

bestfitting

**bestfitting**

Joined 9 years ago · last seen 2 months ago

Competitions Grandmaster  
34 of 203,493

About Competitions (57) Datasets (3) Code (3) Discussion (270) Followers (15475)

### Kaggle Achievements

Follow Contact

Achievement	Rank	Medals
Competitions Grandmaster	34 of 203,493	42 Gold, 13 Silver, 2 Bronze
Datasets Contributor	1 of 28,364	1 Gold
Notebooks Contributor	856 of 28,364	1 Gold, 2 Silver
Discussions Expert	23 of 28,364	17 Gold, 16 Silver, 141 Bronze



Figure source: <https://www.kaggle.com/bestfitting>

# Outline of tasks

---

Platform	Competition Name	Data Type	Task Type	Why special? (Difficulty)
Kaggle	<a href="#">Dog Breed Identification</a>	Image	Image Classification	120 classes
	<a href="#">Plant Pathology 2020 - FGVC7</a>	Image	Image Classification	Class imbalance
	<a href="#">Natural Language Processing with Disaster Tweets</a>	Text	Text Classification	NLP






# Dog Breed Identification

- <https://www.kaggle.com/competitions/dog-breed-identification/overview>

000bec180eb18c7604dcecc8fe0dba07.jpg (54.78 kB)



**Data Explorer**

750.43 MB


- test
- train
- ▮ labels.csv
- ▮ sample\_submission.csv

---

**Summary**

- 20.6k files
- 123 columns

---

 **Download All**



# Dog Breed Identification

affenpinscher  
afghan\_hound  
african\_hunting\_dog  
airedale  
american\_staffordshire\_terrier  
appenzeller  
australian\_terrier  
basenji  
basset  
beagle  
bedlington\_terrier  
bernese\_mountain\_dog  
black-and-tan\_coonhound  
blenheim\_spaniel  
bloodhound  
bluetick  
border\_collie  
border\_terrier  
borzoi  
boston\_bull  
bouvier\_des\_flandres  
boxer  
brabancon\_griffon  
briard  
brittany\_spaniel

⋮

000bec180eb18c7604dcecc8fe0dba07.jpg (54.78 kB)



labels.csv (482.06 kB)

Detail	Compact	Column
⌵ id	≡	⌵ breed
<b>10222</b> unique values		scottish_deerhound 1% maltese_dog 1% Other (9979) 98%
000bec180eb18c7604dcecc8fe0dba07		boston_bull
001513dfcb2ffafc82cccf4d8bbaba97		dingo
001cdf01b096e06d78e9e5112d419397		pekinese
00214f311d5d2247d5dfe4fe24b2303d		bluetick
0021f9ceb3235effd7fcde7f7538ed62		golden_retriever
002211c81b498ef88e1b40b9abf84e1d		bedlington_terrier
00290d3e1fdd27226ba27a8ce248ce85		bedlington_terrier



# Dog Breed Identification (evaluation)

---

## Multi Class Log Loss:

$$\text{loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(\hat{y}_{ij})$$

- 值越小越好
  - $N$ : 資料數量 (test set 的數量)
  - $M$ : 類別數量 (120)
  - $y_{ij}$ : 真實的答案
  - $\hat{y}_{ij}$ : 模型預測的答案


Cross-entropy:  $\mathcal{L}_i = -\log P(Y = y_i | X = x_i)$




# Plant Pathology 2020 - FGVC7

- <https://www.kaggle.com/competitions/plant-pathology-2020-fgvc7/data?select=images>


**images** (3642 files)




Test\_0.jpg  
232.54 kB




Test\_1.jpg  
301.02 kB




Test\_10.jpg  
250.57 kB



Test\_100.jpg  
281.3 kB



Test\_1000.jpg  
243.87 kB



Test\_1001.jpg  
230.99 kB

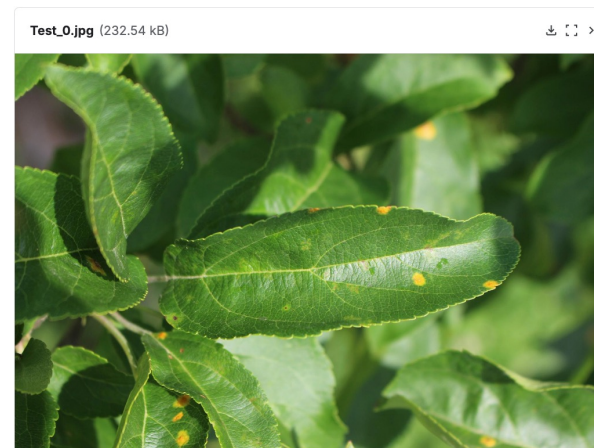
**Data Explorer**  
823.79 MB

- images
  - sample\_submission.csv
  - test.csv
  - train.csv



# Plant Pathology 2020 - FGVC7

- 每張影像有4個類別
  - Healthy
  - Multiple diseases (多重疾病)
  - Rust (鏽病)
  - Scab (疥瘡病)



**train.csv** (33.53 kB) 📄 🔍 >

Detail Compact Column 5 of 5 columns ▾

▲ image_id ▾	# healthy ▾	# multiple_diseases ▾	# rust ▾	# scab ▾
<b>1821</b> unique values	 0 1	 0 1	 0 1	 0 1
Train_0	0	0	0	1



# Plant Pathology 2020- FGVC7 (Evaluation)

---

## Evaluation

Submissions are evaluated on **mean column-wise ROC AUC**. In other words, the score is the average of the individual AUCs of each predicted column.

針對每個 label 的模型預測各算一次 AUC (Area Under the Curve) 之後，取平均





# ROC Curve

(Useful!!) <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=zh-tw>

ROC: Receiver operating characteristic curve (接收者操作特徵曲線)

AUC: ROC 的底面下面積 (越大越好)

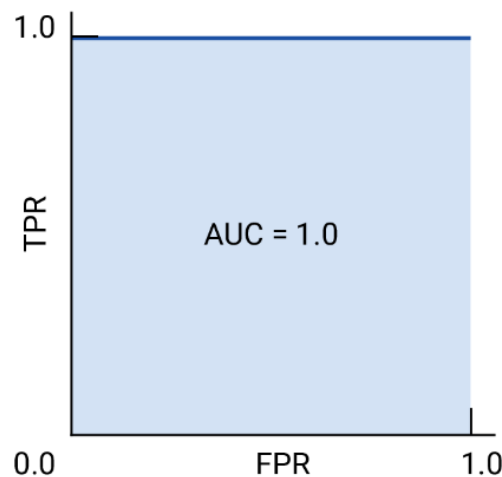


圖 1：ROC 和 AUC，這是一個完美假設的模型。

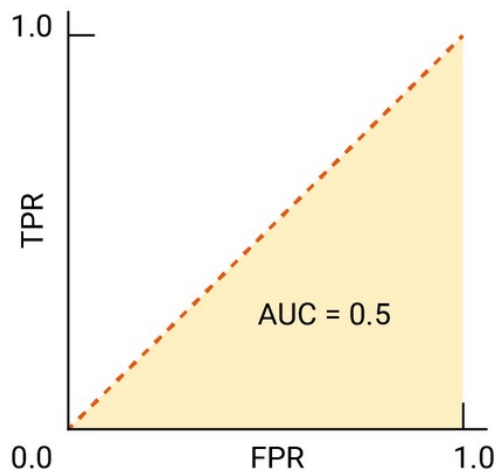


圖 2 ROC 和 AUC 完全隨機猜測。

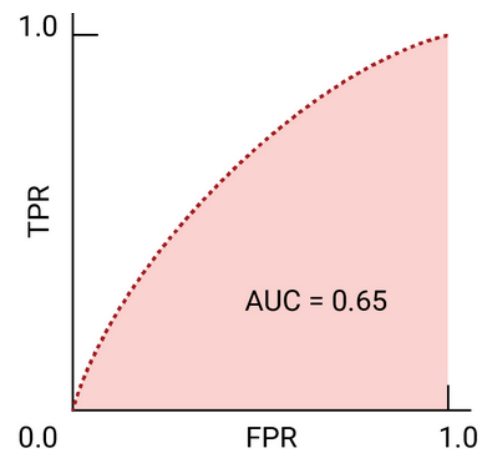
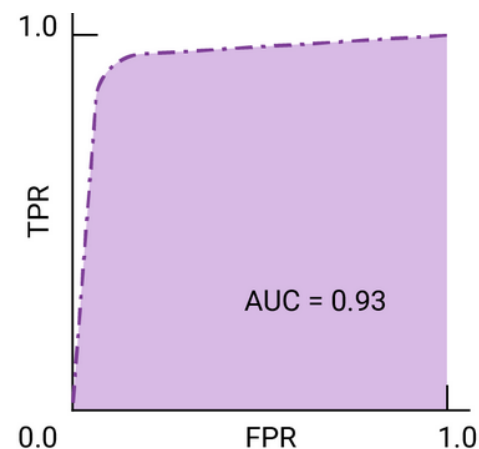
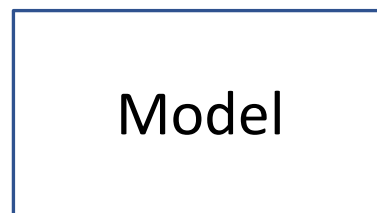
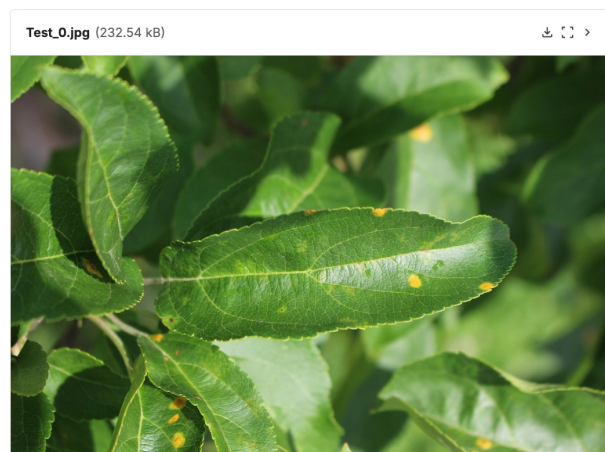


圖 3. 包含兩個假想模型的 ROC 和 AUC。圖表上的而 AUC 越多代表這兩個模型越好



# 模型預測門檻 (Threshold)



模型輸出

機率 = 0.55

**Rust**

答案

1

我們應該設多少門檻？

threshold = 0.5 -> 模型預測為1，正確數量+1

threshold = 0.6 -> 模型預測為0，正確數量-1





# Confusion Matrix (混淆矩陣) for ROC

	Actually positive	Actually negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

- $TP / (TP + FN)$ : True Positive Rate (TPR)
  - 又稱作 Recall
- $FP / (FP + TN)$ : False Positive Rate (FPR)



# 完美模型的條件

- 分類結果完全正確
  - 沒有誤判
    - No False Negatives: TPR 越高越好
    - No False Positives: FPR 越低越好
- 完美模型中存在一個「理想 threshold 區間」，使我們能同時得到  $TPR = 1$  且  $FPR = 0$

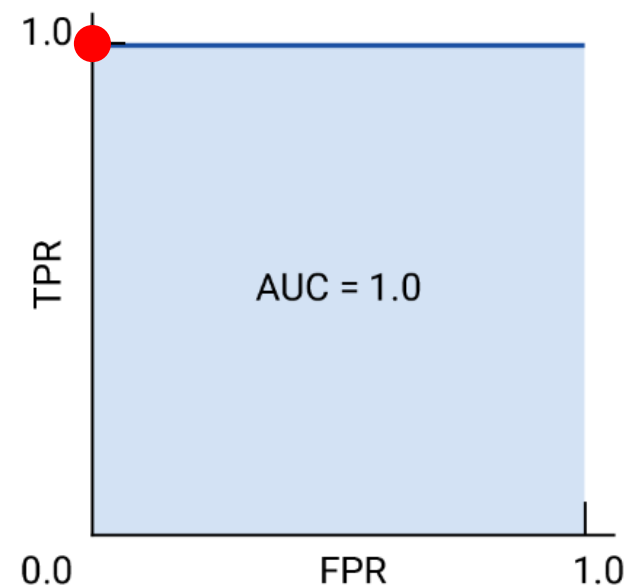


圖1：ROC 和 AUC，這是一個完美假設的模型。



# Confusion Matrix for General Uses

	Actually positive	Actually negative
Predicted positive	True positive (TP) ↑	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN) ↑

- 斜對角數值越大越好
  - TP 和 TN 越大越好
- $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$ 
  - 模型預測的TP比例
- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$ 
  - True Positive Rate (TPR)



# roc\_curve using scikit-learn

---

- [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc\\_curve.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html)

```
import numpy as np
from sklearn import metrics

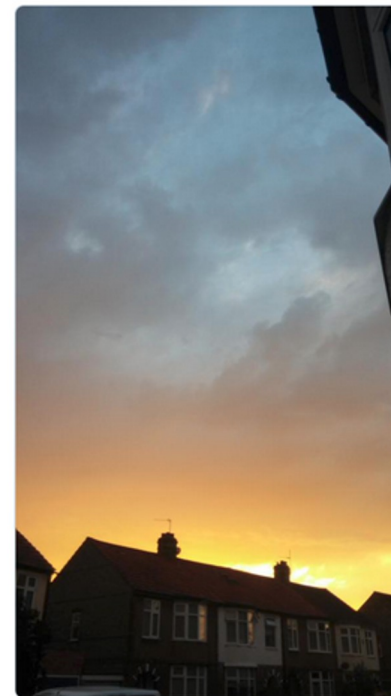
y = np.array([1, 1, 2, 2])
scores = np.array([0.1, 0.4, 0.35, 0.8])
fpr, tpr, thresholds = metrics.roc_curve(y, scores, pos_label=2)
```



# Natural Language Processing with Disaster Tweets

- You are predicting whether a given tweet is about a real disaster or not. If so, predict a 1.
  - If not, predict a 0.
- 資料範例 ➡ (但任務中只需要進行文字分類)

Anna K  
@AnyOtherAnnaK  
On plus side LOOK AT THE SKY LAST NIGHT IT WAS ABLAZE



12:43 AM - Aug 6, 2015 - Twitter for Android

The author explicitly uses the word “ABLAZE” but means it metaphorically. This is clear to a human right away, especially with the visual aid. But it’s less clear to a machine.

In this competition, you’re challenged to build a machine learning model that predicts which Tweets are about real disasters and which one’s aren’t. You’ll have access to a dataset of 10,000 tweets that were hand classified. If this is your first time working on an NLP problem, we’ve created a [quick tutorial](#) to get you up and running.

Disclaimer: The dataset for this competition contains text that may be considered profane, vulgar, or offensive.

## Acknowledgments

This dataset was created by the company figure-eight and originally shared on their

[‘Data For Everyone’ website here.](#)

Tweet source: <https://twitter.com/AnyOtherAnnaK/status/629195955506708480>





# Natural Language Processing with Disaster Tweets

- Data structure and example

**train.csv** (987.71 kB) ↓ [ ] > 1.43 MB

**Detail** Compact Column 5 of 5 columns ▼

id	keyword	location	text	target
 1 10.9k	<b>222</b> unique values	<b>[null]</b> 33% USA 1% Other (4976) 65%	<b>7503</b> unique values	 0 1
48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze <a href="http://t.co/1HYXE0HY6C">http://t.co/1HYXE0HY6C</a>	1
49	ablaze	Est. September 2012 - Bristol	We always try to bring the heavy. #metal #RT <a href="http://t.co/YAo1e0xngw">http://t.co/YAo1e0xngw</a>	0
50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set ablaze in Aba. <a href="http://t.co/2nndBGwyEi">http://t.co/2nndBGwyEi</a>	1

sample\_submission.csv  
test.csv  
train.csv

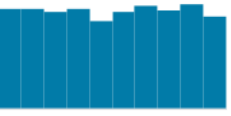


# Natural Language Processing with Disaster Tweets

- How to submit?

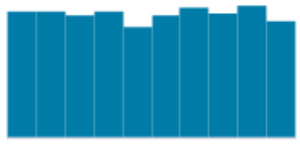

test.csv (420.78 kB)

Detail Compact Column

id	keyword	location	text
 010.9k	<b>222</b> unique values	<b>[null]</b> 34% New York 1% Other (2120) 65%	<b>3243</b> unique values
0			Just happened a terrible car crash
2			Heard about #earthquake is different cities, stay safe everyone.
3			there is a forest fire at spot pond, geese are fleeing across the street, I cannot save them all
9			Apocalypse lighting. #Spokane #wildfires
11			Typhoon Soudelor kills 28 in China and Taiwan

sample\_submission.csv (22.75 kB)

Detail Compact Column

id	# target
 010.9k	 00
0	0
2	0
3	0
9	0
11	0



# Natural Language Processing with Disaster Tweets

- 分數計算? 採用 F1-score

	Actually positive	Actually negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

- $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$ 
  - 模型預測的TP比例
- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$ 
  - True Positive Rate (TPR)
- $\text{F1-score (廣義)} = 2(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$





# Candidates

---

- Google Open Images Object Detection RVC 2020 edition
  - <https://www.kaggle.com/competitions/open-images-object-detection-rvc-2020>
- NBME - Score Clinical Patient Notes (Code competition)
  - <https://www.kaggle.com/competitions/nbme-score-clinical-patient-notes/overview>
- Google QUEST Q&A Labeling (Notebook competition)
  - <https://www.kaggle.com/competitions/google-quest-challenge>



# Project checkpoints (暫定)

---

- Week 9: 確定各組的題目
- Week 11: 進度報告 PPT (5 pages)
- Week 13: 進度報告 PPT (5+5 pages), Presentations (selected teams)
- Week 15 – Week 16: Final presentations for all teams (maybe poster)
- Week 16 結束前: 繳交書面報告以及程式碼



# 期末 Project 規定 (暫定)

---

- 需要上傳 Kaggle Leaderboard
  - 這些任務都僅需要上傳 predictions 到 Kaggle 即可
- 每次報告都需要列出每位組員的貢獻內容，以及組員間的工作比重 (%)



# Thank you!

Instructor: 林英嘉

 yjlin@cgu.edu.tw

TA: 林君襄

 becky890926@gmail.com