

認識 AI 是甚麼

整理人：陳祥輝

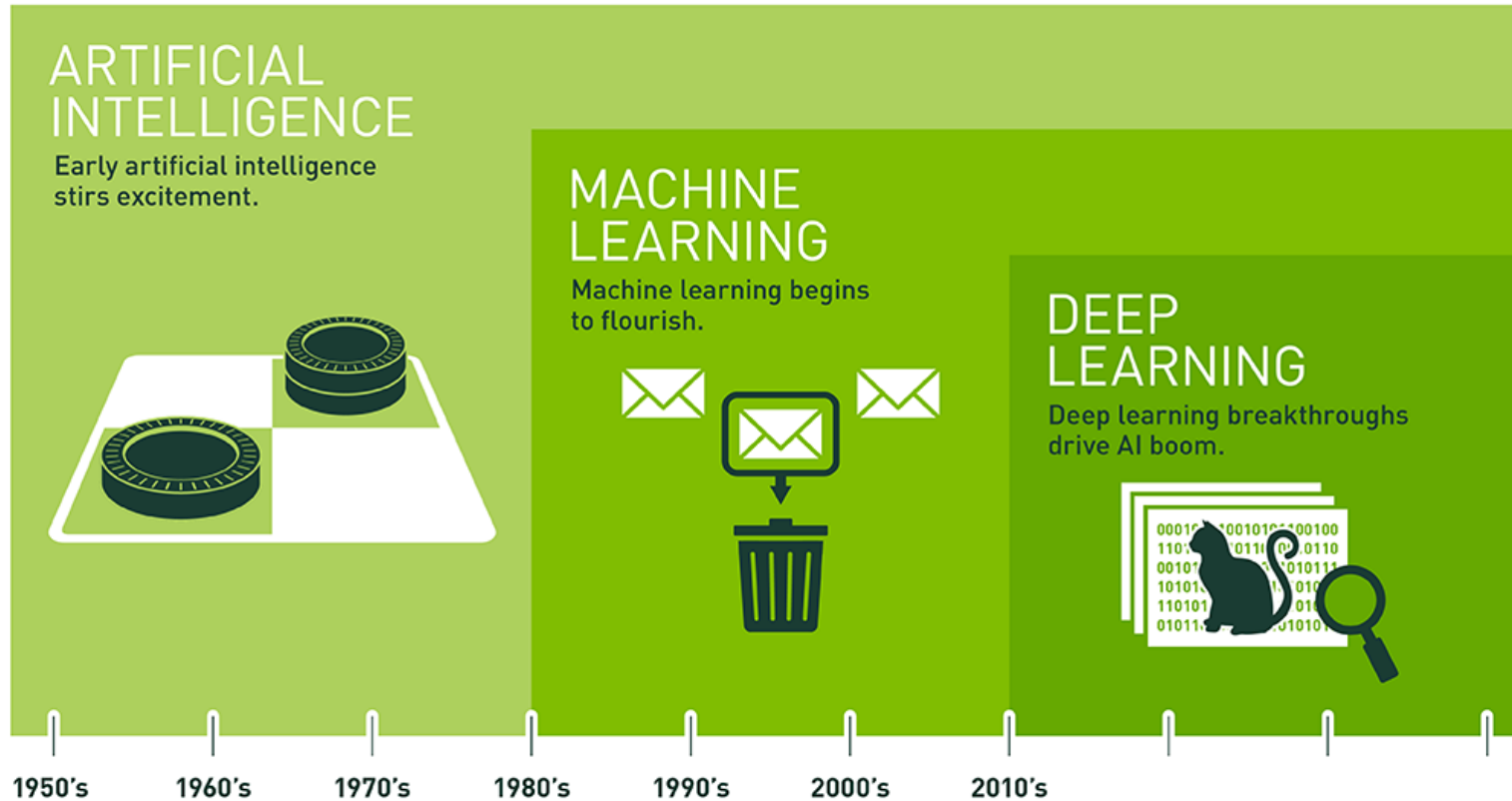
Email : HsiangHui.Chen@gmail.com

日期：2/28/2025

數據分析的種類

- **Business Intelligence (BI, 商業智慧)**
 - 多維度分析、視覺化分析
 - Excel樞紐分析、Power BI (Desktop)、Tableau、IBM Cognos、SAS VA
- **Artificial Intelligence (AI, 人工智慧)**
 - Machine Learning (ML)
 - Deep Learning (DL)
- **Statistics (統計分析)**
 - 敘述性統計(描述性統計)
 - 推論統計

Artificial Intelligence (AI)

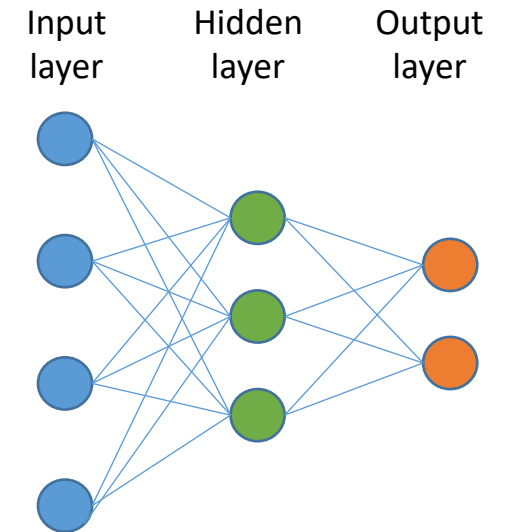


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

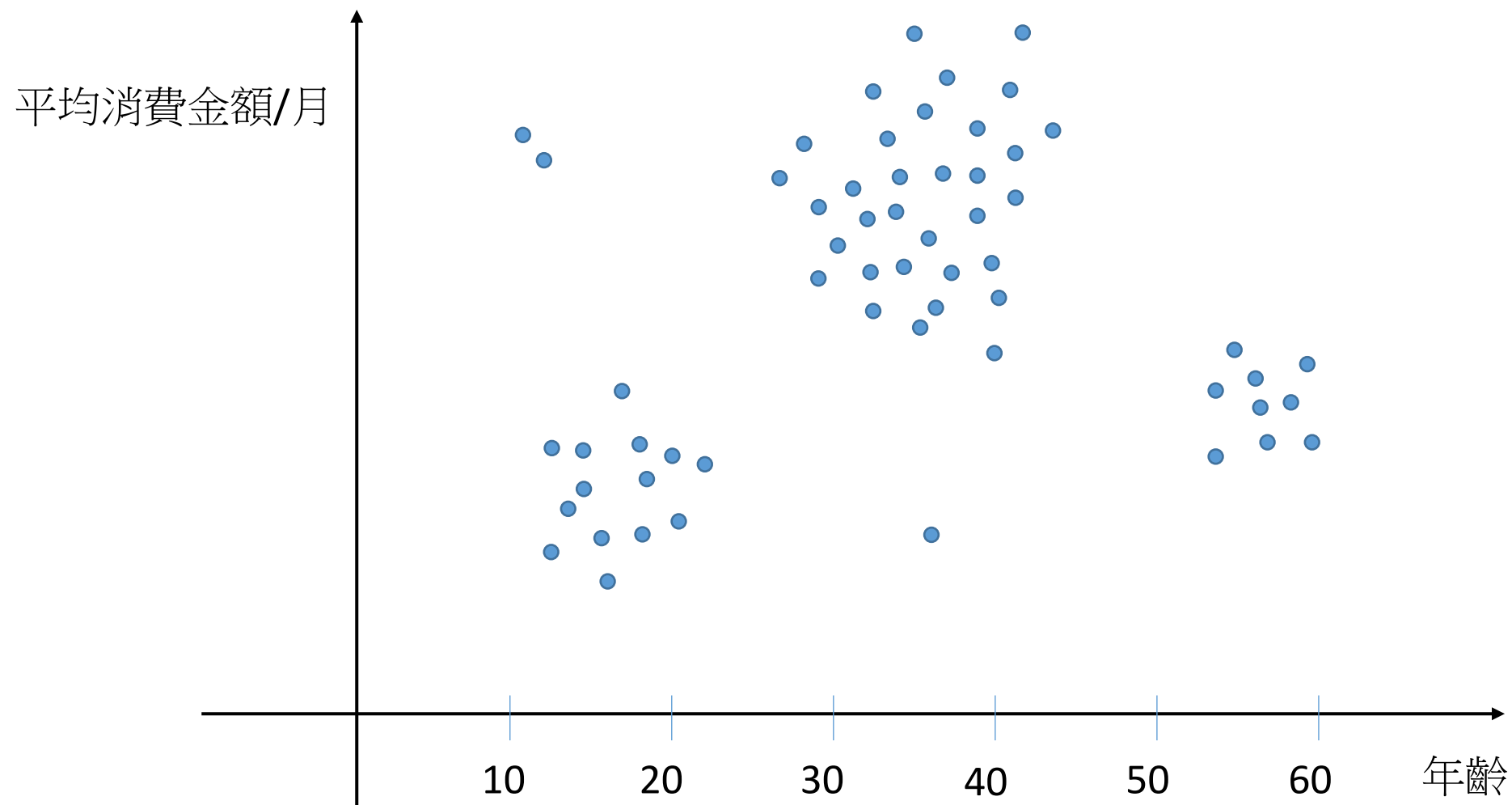
[資料來源] <https://blogs.nvidia.com.tw/2016/07/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

Overview of Machine Learning

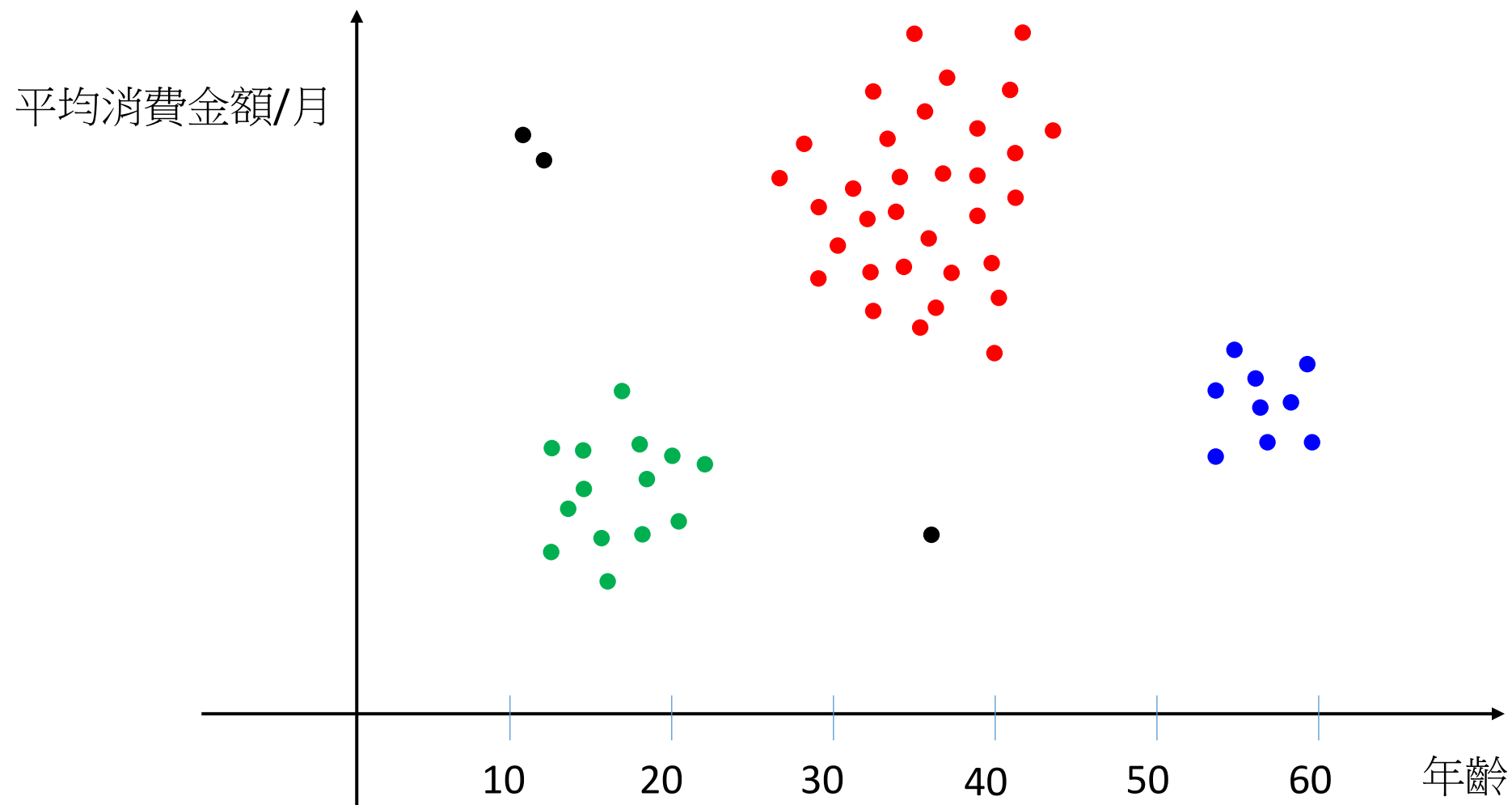
- Supervised Learning (監督式學習)
 - Classification (分類)
 - Regression (迴歸)
- Unsupervised Learning (非監督式學習)
 - Clustering Analysis (集群分析)
- Reinforcement Learning (增強式學習)
- Time Series (時間序列)
- Neural Network
 - Single Layer Perceptron / Multi Layer Perceptron (MLP)
 - Deep Learning



非監督式學習



非監督式學習





假如您是一部機器，我要訓練您學習



貓

狗



貓



狗



貓

貓



貓

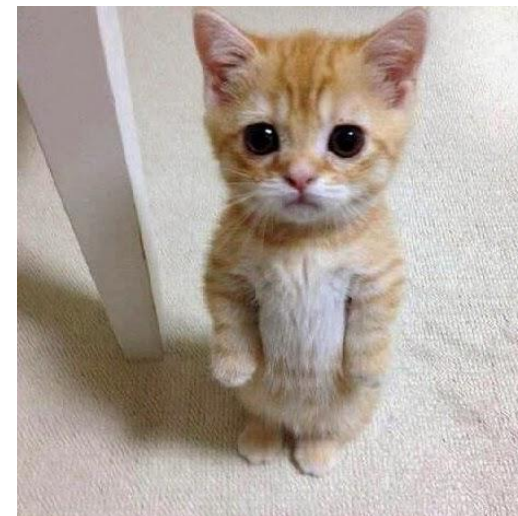


狗

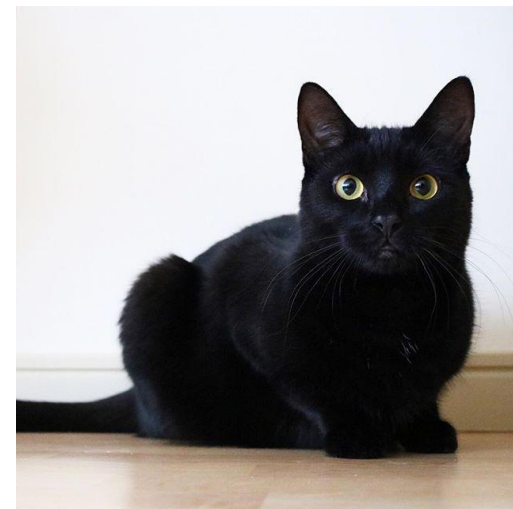
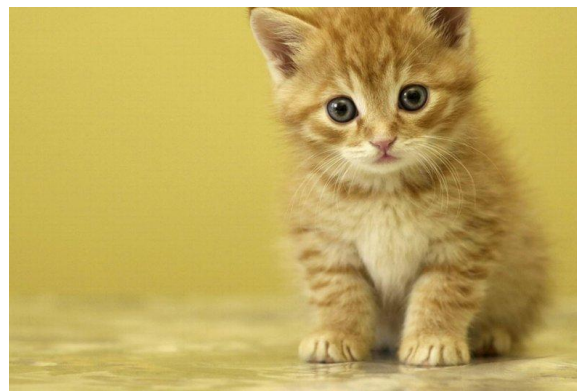
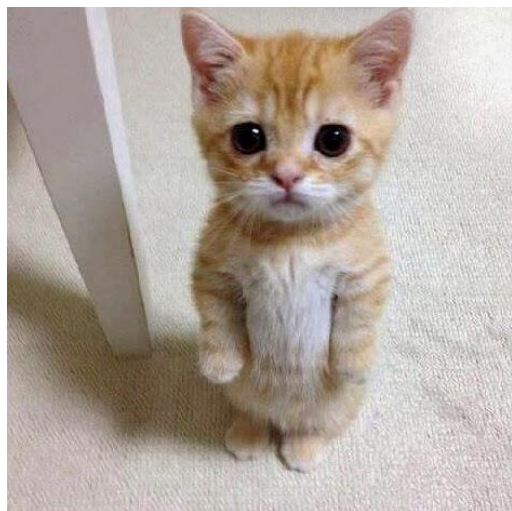
請機器預測



Y，我咧？

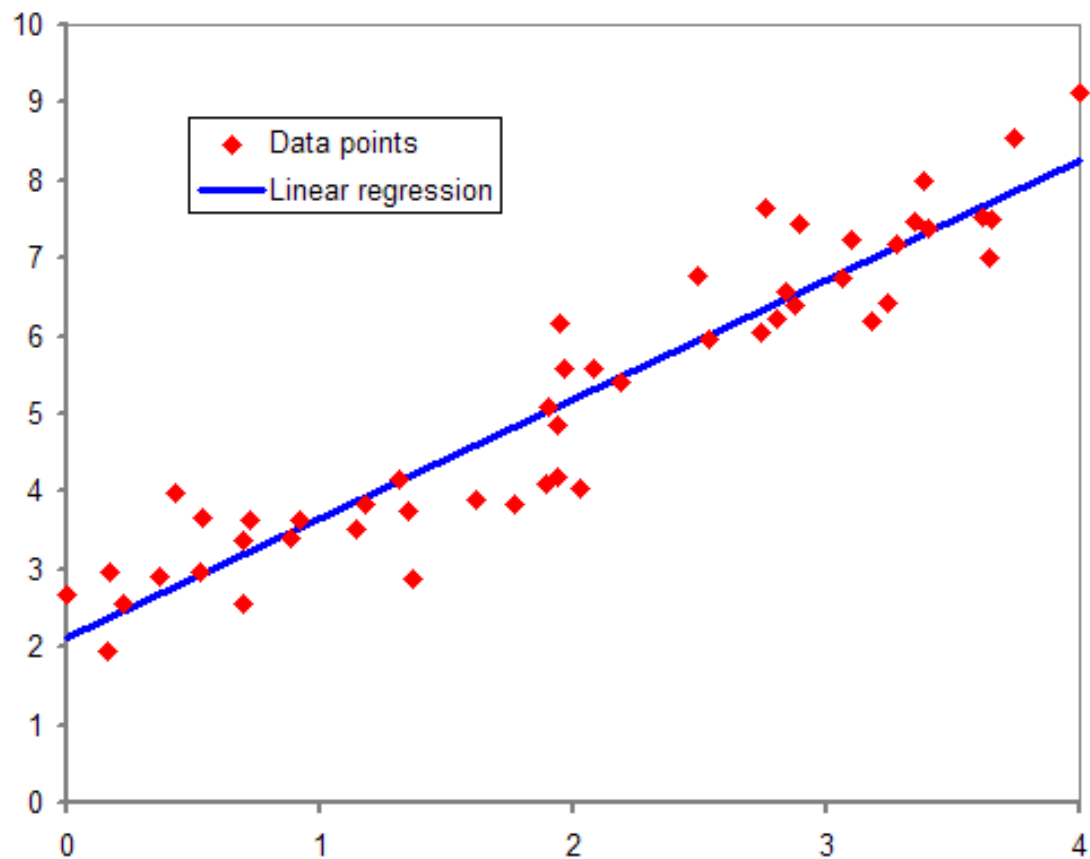


請機器幫我分群，不限群數，不一定等分



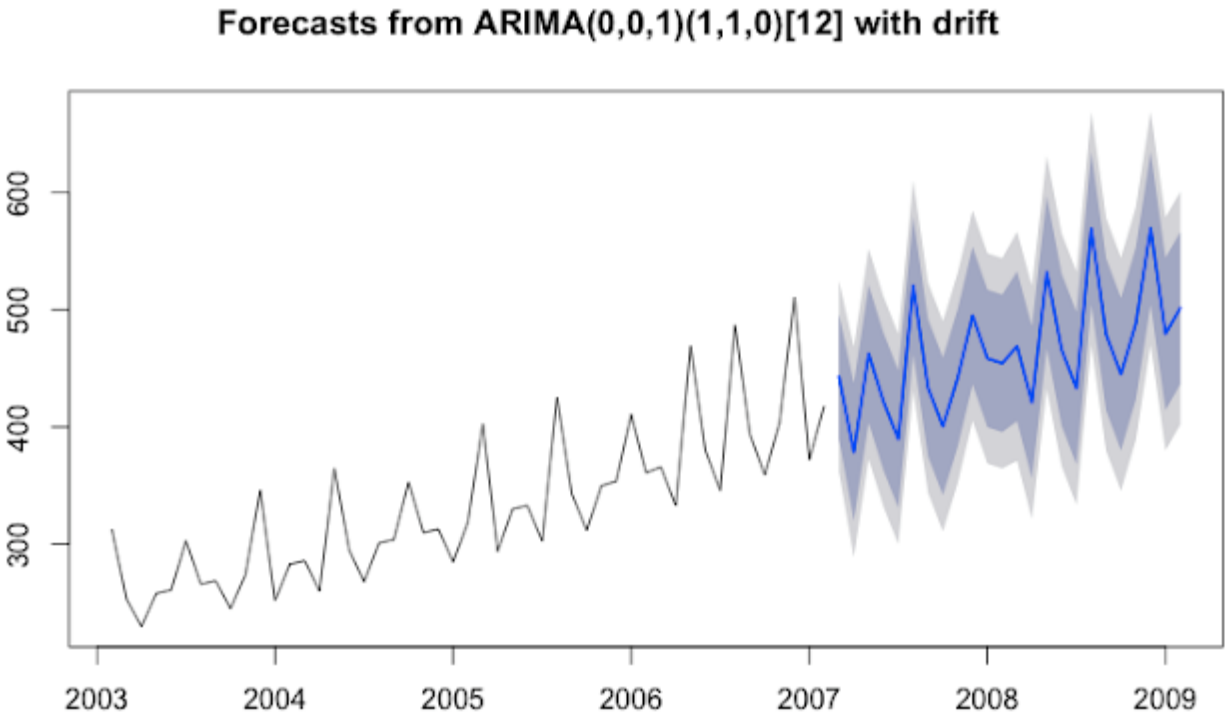
線性迴歸 (Linear Regression)

$$y = \beta_0 + \beta_1 x$$



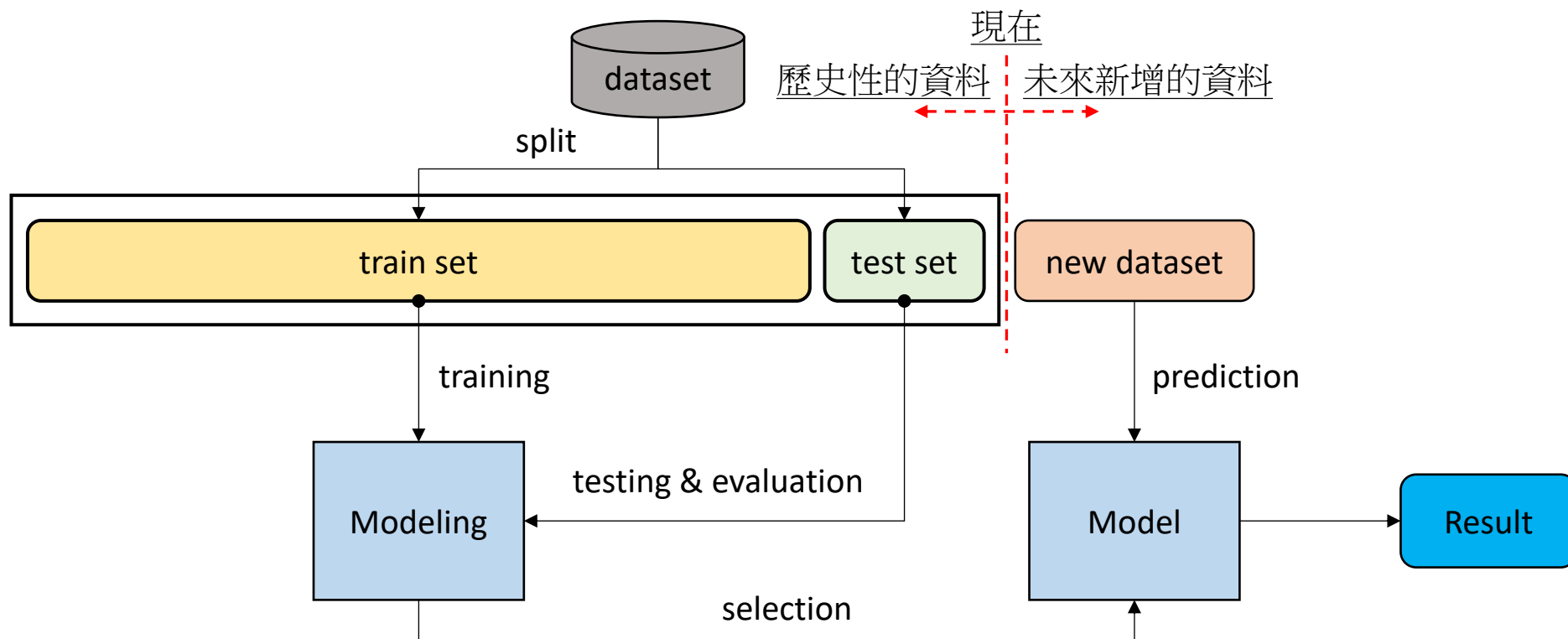
speed	dist
<dbl>	<dbl>
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26
10	34
11	17
11	28

Time Series (時間序列)



'GOOG'	開盤價	最高價	最低價	收盤價	成交量	還原權值
	GOOG.Open	GOOG.High	GOOG.Low	GOOG.Close	GOOG.Volume	GOOG.Adjusted
2019-11-01	1265.00	1274.620	1260.500	1273.74	1670100	1273.74
2019-11-04	1276.45	1294.130	1276.355	1291.37	1501000	1291.37
2019-11-05	1292.89	1298.930	1291.229	1292.03	1282700	1292.03
2019-11-06	1289.46	1293.730	1282.500	1291.80	1152700	1291.80
2019-11-07	1294.28	1323.740	1294.245	1308.86	2030000	1308.86
2019-11-08	1305.28	1318.000	1304.365	1311.37	1251400	1311.37
2019-11-11	1303.18	1306.425	1297.410	1299.19	1011900	1299.19
2019-11-12	1300.00	1310.000	1295.770	1298.80	1085900	1298.80
2019-11-13	1294.07	1304.300	1293.510	1298.00	826700	1298.00
2019-11-14	1297.50	1317.000	1295.650	1311.46	1193500	1311.46
2019-11-15	1318.94	1334.880	1314.280	1334.87	1782600	1334.87
2019-11-18	1332.22	1335.529	1317.500	1320.70	1487400	1320.70
2019-11-19	1327.70	1327.700	1312.800	1315.46	1269200	1315.46
2019-11-20	1311.74	1315.000	1291.150	1303.05	1308600	1303.05
2019-11-21	1301.48	1312.590	1293.000	1301.35	995500	1301.35
2019-11-22	1305.62	1308.730	1291.410	1295.34	1385700	1295.34
2019-11-25	1299.18	1311.310	1298.130	1306.69	1036200	1306.69
2019-11-26	1309.86	1314.800	1305.090	1313.55	1069700	1313.55
2019-11-27	1315.00	1318.360	1309.630	1312.99	995600	1312.99
2019-11-29	1307.12	1310.205	1303.970	1304.96	587000	1304.96

Training, Testing, Prediction

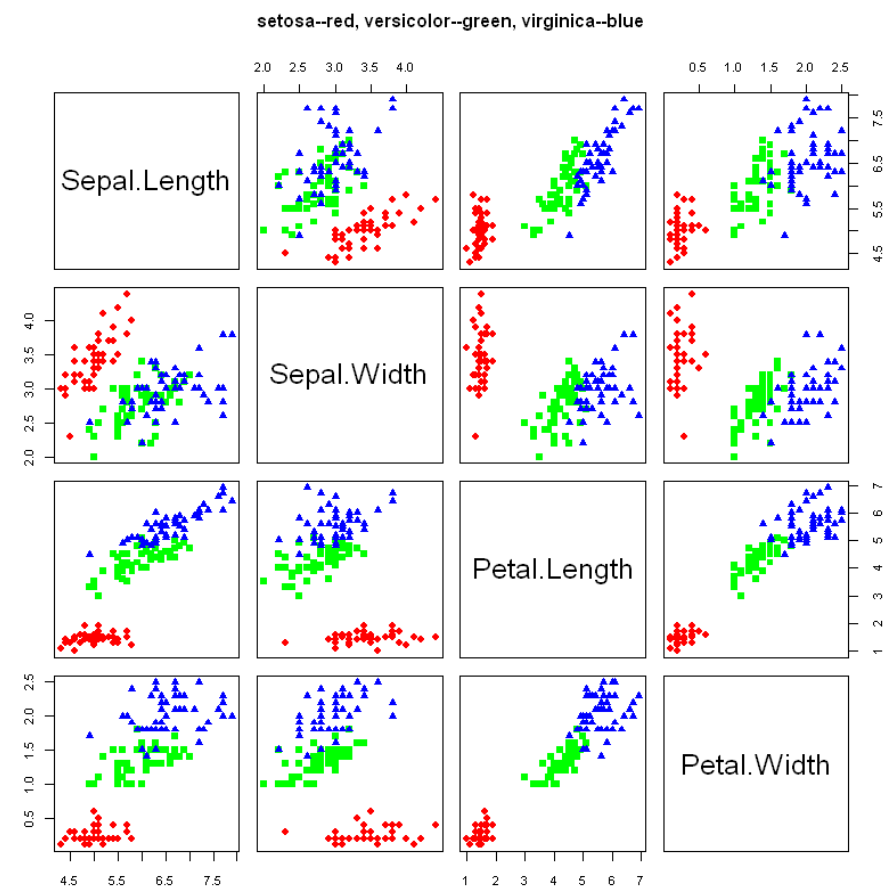


資料集範例

花萼長度 花萼寬度 花瓣寬度 花瓣長度 種類

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa

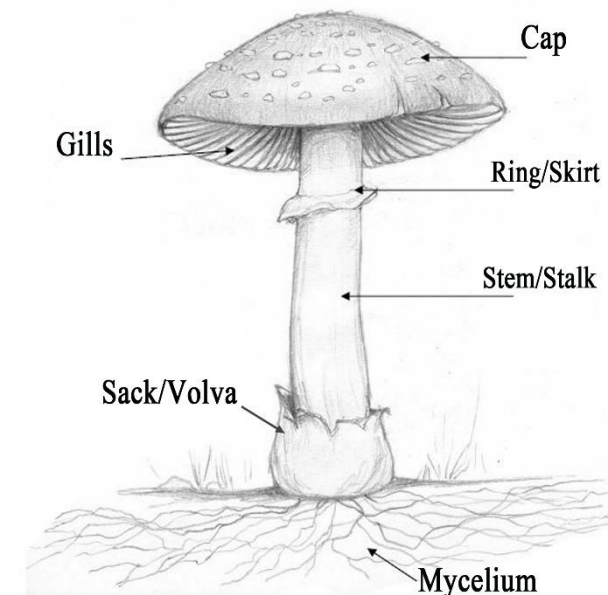
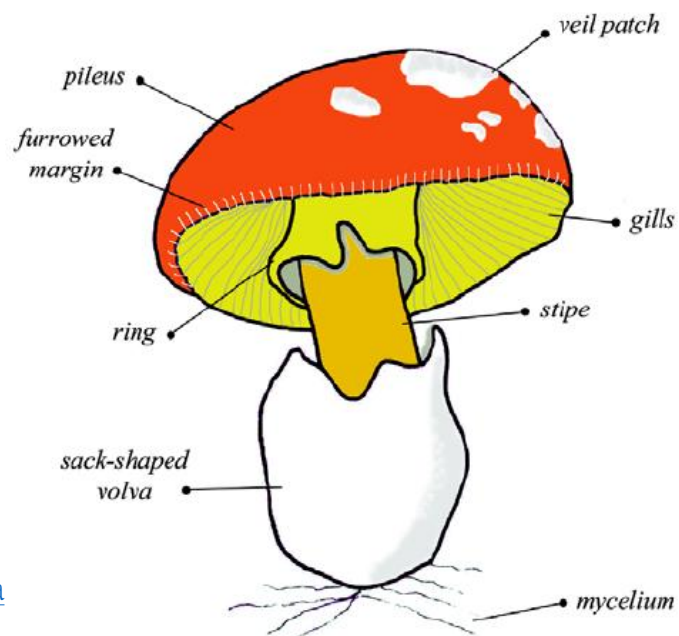
Showing 1 to 15 of 150 entries



欄位名稱	說明	備註
Sepal.Length	花萼長度	
Sepal.Width	花萼寬度	
Petal.Length	花瓣長度	
Petal.Width	花瓣寬度	
Species	種類	setosa versicolor(雲芝) virginica(錦葵)

Mushroom(蘑菇)

- class : **edible(可食用)=e, poisonous(有毒)=p**
- cap-shape : bell=b, conical=c, convex=x ,flat=f, knobbed=k, sunken=s
- cap-surface : fibrous=f, grooves=g, scaly=y, smooth=s
- cap-color : brown=n, buff=b, cinnamon=c, gray=g, green=r ,pink=p ,purple=u ,red=e, white=w, yellow=y
- bruises(傷痕) : bruises=t, no=f
- odor(氣味) : almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s
- gill-attachment : attached=a, descending=d, free=f, notched=n
- gill-spacing : close=c, crowded=w, distant=d
- gill-size : broad=b, narrow=n
- gill-color : black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
- stalk-shape
- stalk-root
- stalk-surface-above-ring
- stalk-surface-below-ring
- stalk-color-above-ring
- stalk-color-below-ring
- veil-type
- veil-color
- ring-number
- ring-type
- spore-print-color (孢子印色)
- population (族群)
- habitat (棲息地)



Red Wine Quality Data Set

固定酸	揮發酸性	檸檬酸	殘留糖	氯化物	游離二氧化硫	總二氧化硫	密度	PH值	硫酸鹽	酒精	品質
7.4	0.70	0.00	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.8	0.88	0.00	2.6	0.098	25	67	0.9968	3.20	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15	54	0.9970	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17	60	0.9980	3.16	0.58	9.8	6
7.4	0.70	0.00	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.4	0.66	0.00	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5

fixed.acidity	volatile.acidity	citric.acid	residual.sugar	chlorides	free.sulfur.dioxide	total.sulfur.dioxide	density	pH	sulphates	alcohol	quality
7.4	0.70	0.00	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.8	0.88	0.00	2.6	0.098	25	67	0.9968	3.20	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15	54	0.9970	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17	60	0.9980	3.16	0.58	9.8	6
7.4	0.70	0.00	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
7.4	0.66	0.00	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5

心臟疾病的資料集

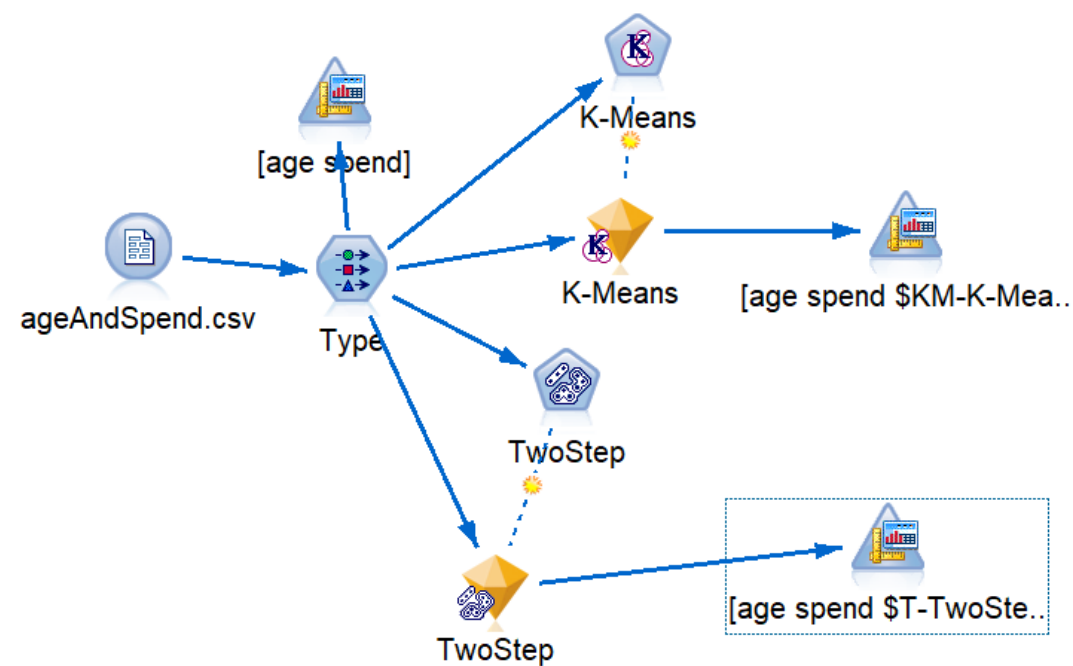
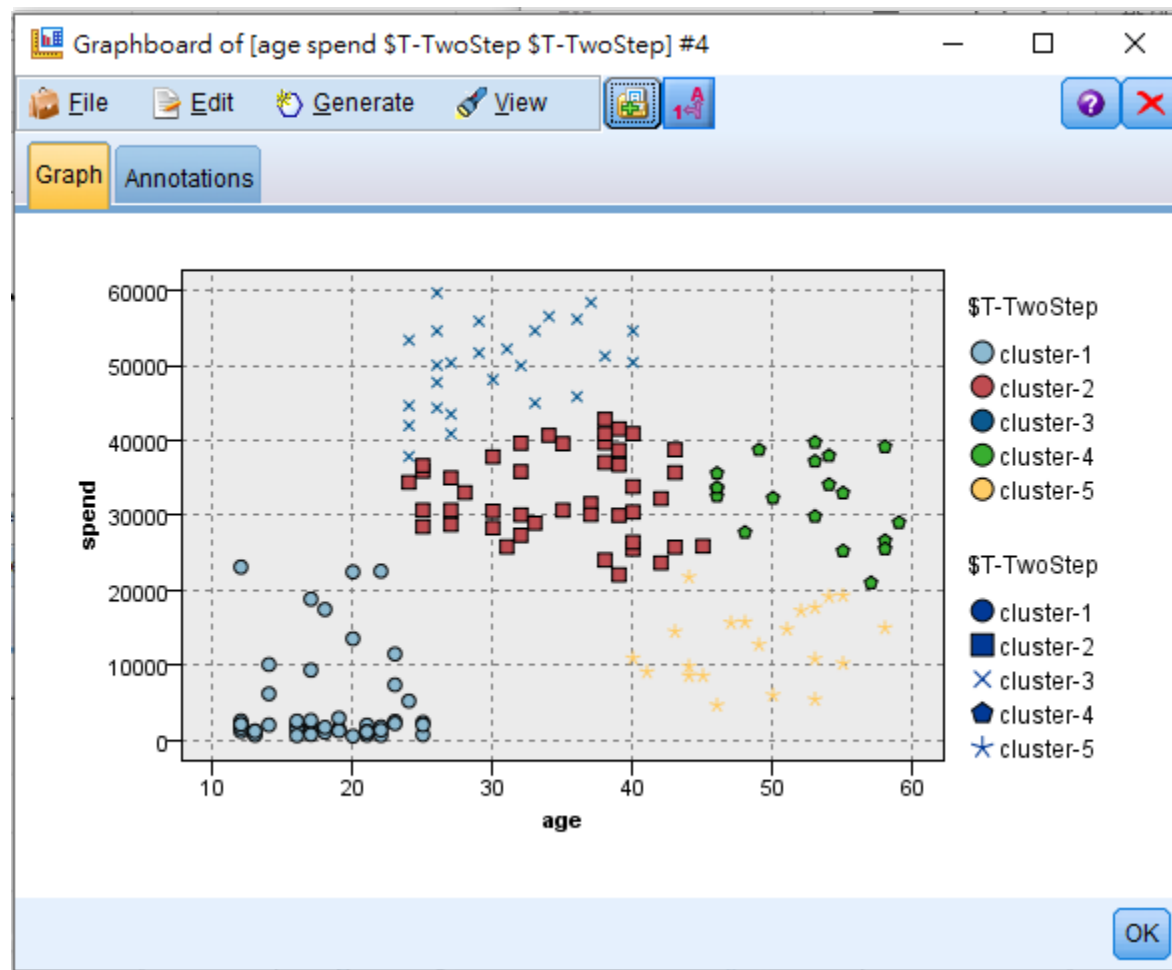
1. age : Age of patient
2. sex : (value 1: Male; value 0 : Female)
3. cp : Chest Pain Type (胸痛類型)
 - value 1: typical type 1 angina (典型的1型心絞痛)
 - value 2: typical type angina (典型的心絞痛)
 - value 3: non-angina pain (非心絞痛)
 - value 4: asymptomatic (無症狀)
4. trestbps : Treat Blood Pressure (mm Hg on admission to the hospital) (治療血壓，入院時)
5. chol : Serum Cholesterol (mg/dl) (血清膽固醇)
6. fbs : Fasting Blood Sugar (空腹血糖)
 - value 1: > 120 mg/dl
 - value 0: < 120 mg/dl
7. restecg : resting electrographic results (靜息電圖結果)
 - value 0: normal
 - value 1: 1 having ST-T wave abnormality
 - value 2: showing probable or definite left ventricular hypertrophy
8. thalach : maximum heart rate achieved (最大心率)
9. exang : exercise induced angina (運動誘發心絞痛)
 - value 1: yes
 - value 0: no
10. oldpeak : ST depression induced by exercise relative to rest (運動引起ST段壓低相對於休息)
11. slope : slope of peak exercise ST (高峰運動ST的斜率)
 - value 1: unsloping
 - value 2: flat
 - value 3: downsloping
12. ca : number of major vessel (value 0 – 3) (主要導管數量)
13. thal : Thal (地中海型貧血)
 - value 3: normal (正常)
 - value 6: fixed defect (固定缺陷)
 - value 7: reversible defect (可逆缺陷)
14. num :
 - value 0 : 沒有心臟病
 - Value 1-4 : 有心臟病

心臟疾病的資料集

該數據庫包含76個屬性，但在所有已發布的實驗中，我們使用其中14個屬性的子集。特別是，克利夫蘭數據庫是ML研究人員迄今為止唯一使用過的數據庫。“目標” num是指患者中存在心臟病。它的整數值從0（無存在）到4.克利夫蘭數據庫的實驗集中於簡單地嘗試區分存在（值1,2,3,4）和不存在（值0）。

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	2
67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1
37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0
41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0
56	1	2	120	236	0	0	178	0	0.8	1	0.0	3.0	0
62	0	4	140	268	0	2	160	0	3.6	3	2.0	3.0	3
57	0	4	120	354	0	0	163	1	0.6	1	0.0	3.0	0
63	1	4	130	254	0	2	147	0	1.4	2	1.0	7.0	2
53	1	4	140	203	1	2	155	1	3.1	3	0.0	7.0	1
57	1	4	140	192	0	0	148	0	0.4	2	0.0	6.0	0

Unsupervised Learning (非監督式學習)



THANK YOU

