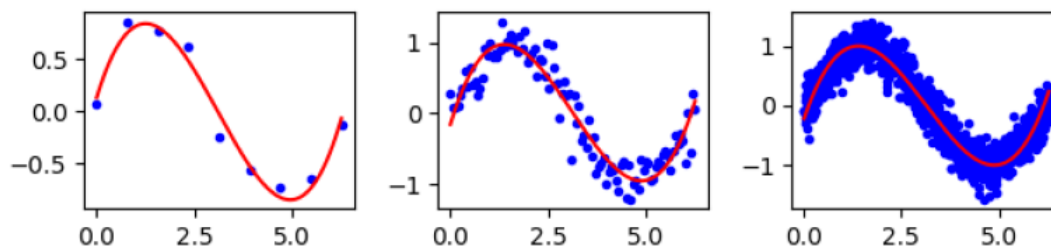


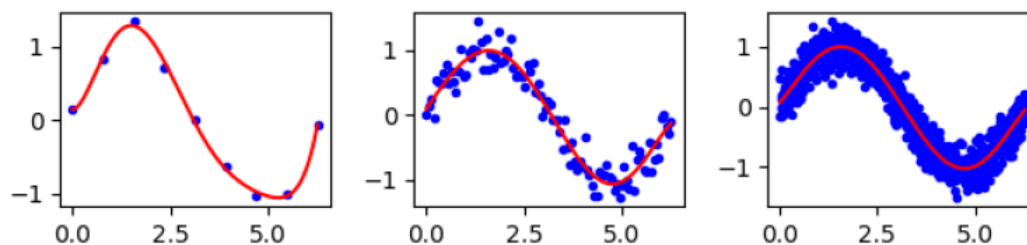
AI 期中報告

7106029070 徐紹恩

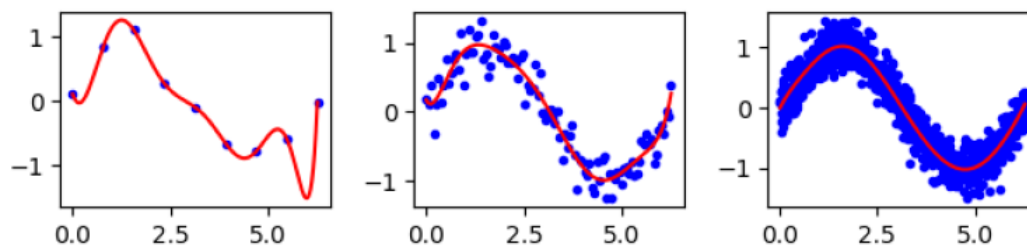
第一周 09/13 回家作業



使用三階多項式做 10 個樣本、100 個樣本、1000 個樣本做多項式凝合。



使用六階多項式做 10 個樣本、100 個樣本、1000 個樣本做多項式凝合。

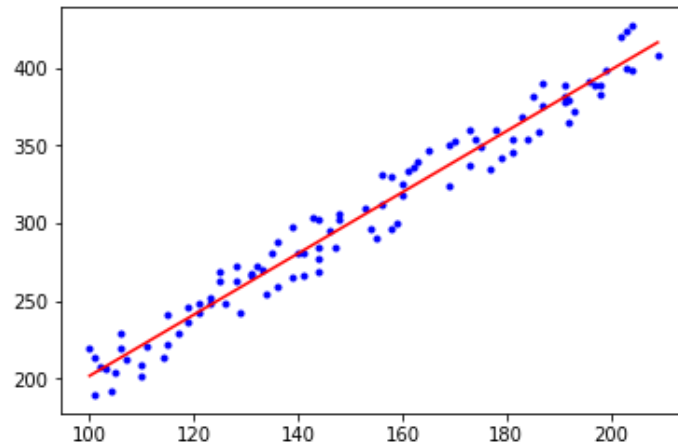


使用九階多項式做 10 個樣本、100 個樣本、1000 個樣本做多項式凝合。

結論：

在樣本數目很少的情況下，盡量使用低階多項是避免出現過凝合的情況，而樣本數夠多的化可以提高多項式的階層。

第二周 09/20 回家作業



Epoch: 50236
loss: 6755.21, Weight: 1.97, Bias: 4.55

使用梯度下降的方式進行多項式擬合

樣本：

隨機生成 100 個線性樣本

收斂條件：

連續 500 個迭代訓練誤差為下降，或是誤差值低於 1 則結束訓練。

訓練結果：

於第 50236 個迭代結束訓練，訓練誤差 6755.21

求出的多項是為：

$$F(X) = 1.97X + 4.55$$

結論：

在訓練過程中發現學習率設太大的話會導致震盪無法收斂。

第三、四周 09/27, 10/04 回家作業

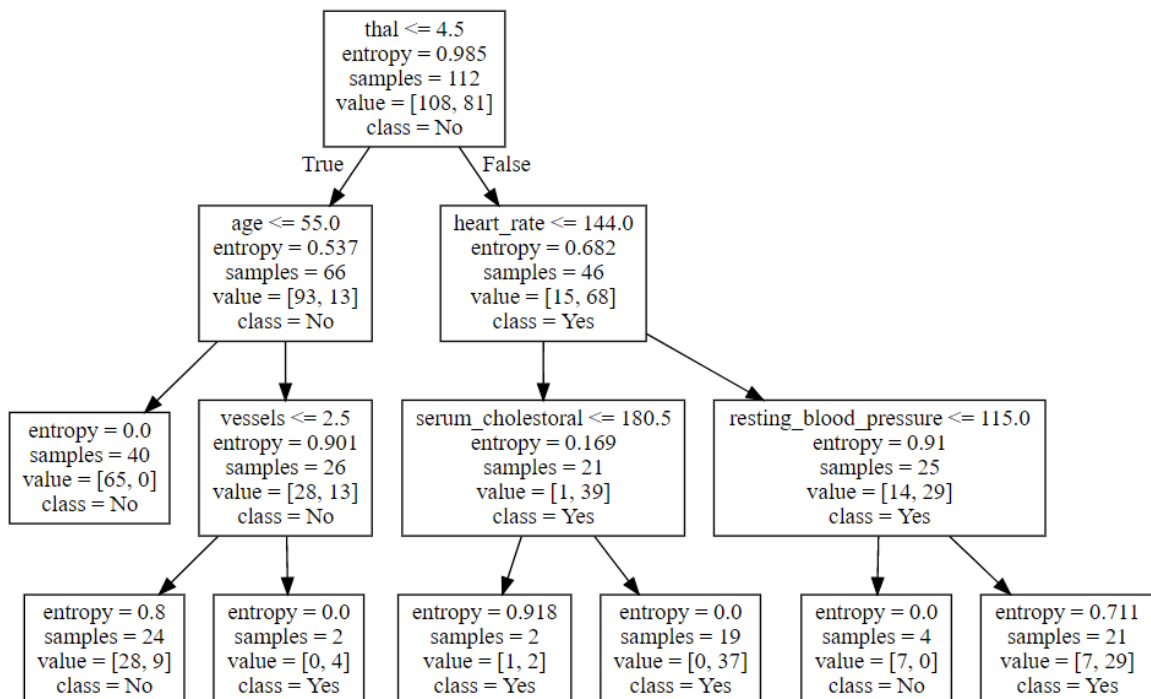
使用 UCI 中的 Statlog (Heart) Data Set 資料集，此資料集為與心臟疾病相關的資料集，為分類任務，以是否有心臟相關疾病為類別進行分類，以下使用相同資料與不同模型進行訓練測試。

資料總筆數：

- 270 筆
- 切分 80/20 做訓練與測試
- 訓練集 189 筆，測試集 81 筆

隨機森林法

使用 Entropy 做為評估標準，樹的最大深度設定為 3。



訓練結果：

訓練集正確率：84.13%

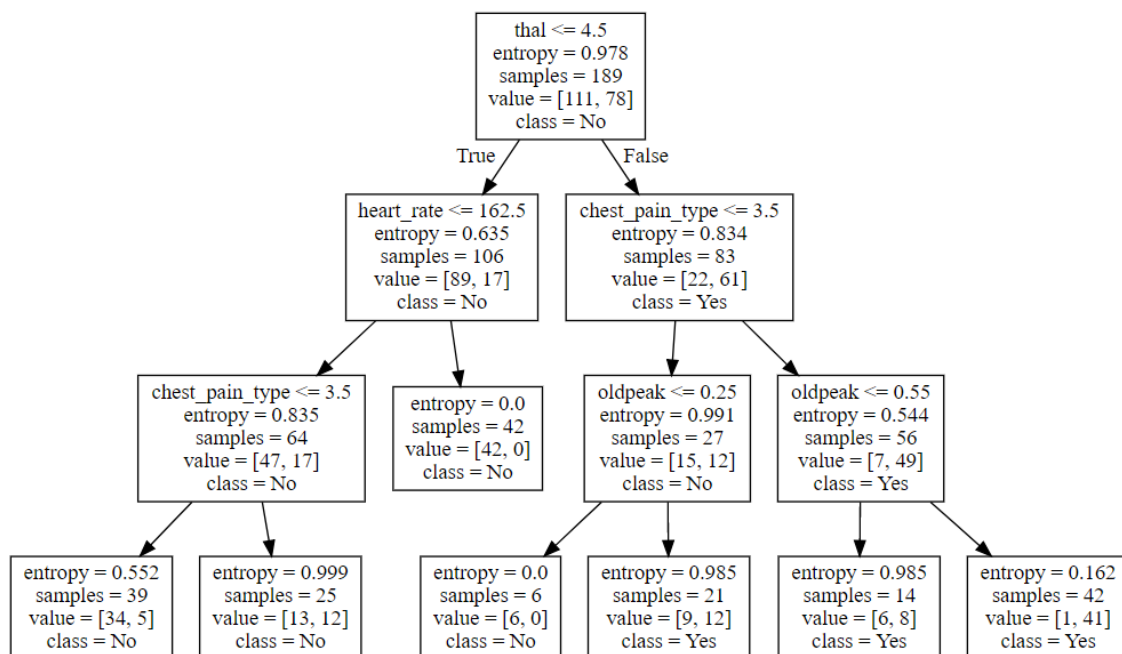
測試集正確率：71.60%

以下為各個特徵在此次分類中所佔的重要性：

	Feature	Feature_importance
12	thal	0.284580
9	oldpeak	0.183576
7	heart_rate	0.143212
11	vessels	0.099113
8	exercise_induced_angina	0.082411
4	serum_cholestorol	0.067933
3	resting_blood_pressure	0.060719
0	age	0.051887
1	sex	0.026568
2	chest_pain_type	0.000000
5	fasting_blood_sugar	0.000000
6	resting_electrocardiographic_results	0.000000
10	ST	0.000000

決策樹

使用 Entropy 做為評估標準，樹的最大深度設定為 3。



訓練結果：

訓練集正確率：82.54%

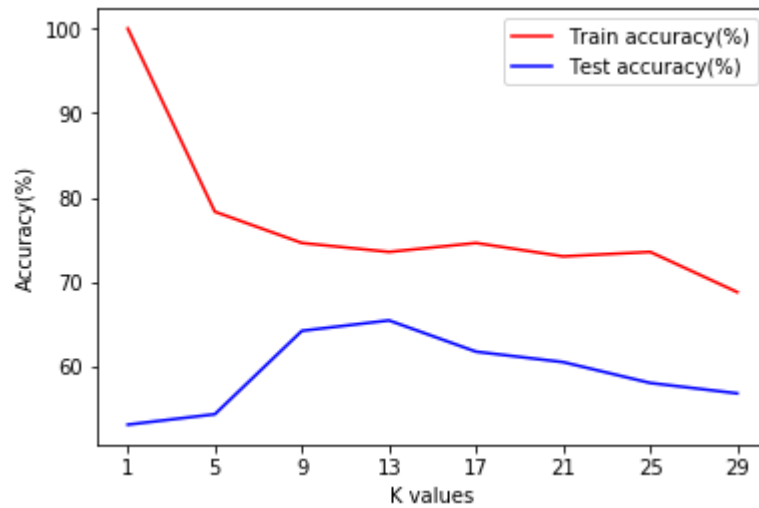
測試集正確率：67.90%

以下為各個特徵在此次分類中所佔的重要性：

	Feature	Feature_importance
12	thal	0.497345
2	chest_pain_type	0.195624
9	oldpeak	0.163891
7	heart_rate	0.143139
0	age	0.000000
1	sex	0.000000
3	resting_blood_pressure	0.000000
4	serum_cholestorol	0.000000
5	fasting_blood_sugar	0.000000
6	resting_electrocardiographic_results	0.000000
8	exercise_induced_angina	0.000000
10	ST	0.000000
11	vessels	0.000000

第五周 10/11 回家作業

使用 KNN 模型做分類，比較各個 K 值所得出的結果。

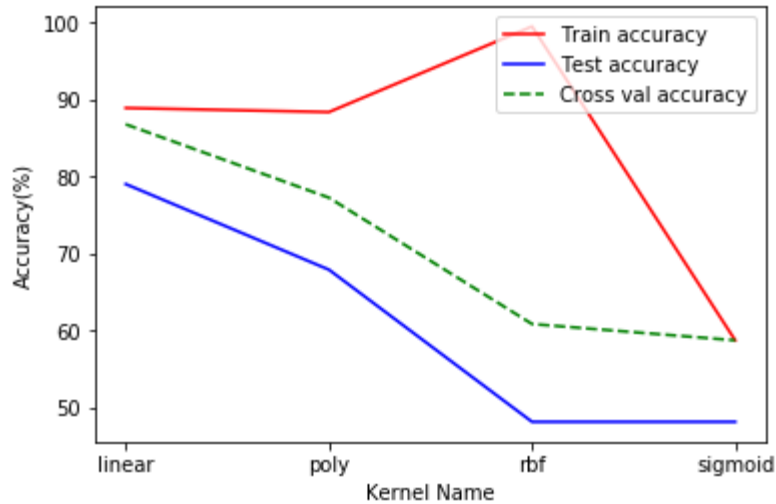


結論：

在本資料集中找出最適的 K 值為 13。

第七周 10/25 回家作業

使用 SVC 模型進行分類，比較不同核心所得出的結果。



以相同的參數 $C=1$ 、 $\text{Gamma}=0.01$ ，不同的核心 linear、poly、rbf、sigmoid 做比較，的出在使用 linear 的核心時做分類出的結果為最好的。