

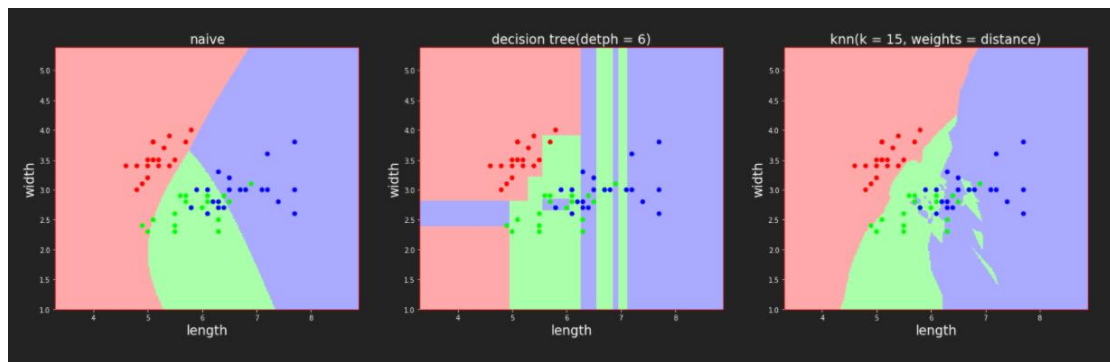
## Homework report(實作機器學習練習) 楊緣智(109522026)

這次是試用看看 sklearn 的套件分別以 Naive Bayesian, knn, decision tree 來訓練鳶尾花的 data 並且做預測。

費雪鳶尾花卉數據集

花萼長度 ◆	花萼寬度 ◆	花瓣長度 ◆	花瓣寬度 ◆	屬種 ◆
5.1	3.5	1.4	0.2	<i>setosa</i>
4.9	3.0	1.4	0.2	<i>setosa</i>
4.7	3.2	1.3	0.2	<i>setosa</i>
4.6	3.1	1.5	0.2	<i>setosa</i>

鳶尾花總共有 *setosa*, *versicolor*, *virginica* 三類。這次是以花萼長度及寬度作為 features 來預測他是哪種“屬種”



訓練完成後以圖中的三色色板(紅、藍、綠)分別代表了所有可能值的預測結果，而紅藍綠的點則代表測試資料的實際值。

```

the accuracy of naive : 0.7666666666666667

naive predict : [0 2 0 1 1 2 0 1 1 0 0 2 2 0 2 0 0 1 1 1 1 1 1 0 0 0 0 1 1 2 1 1 1 1 0 0
 2 1 0 0 0 0 2 2 1 0 2 2 0 2 2 2 2 0 1 1 2 1 0]
real : [0 2 0 1 2 2 1 2 1 0 0 2 1 0 2 0 0 1 1 1 1 1 2 2 0 1 0 0 2 2 2 1 1 2 1 0 0
 2 1 0 0 0 0 1 1 1 0 2 1 0 2 2 2 2 0 1 2 2 1 0]
-----
the accuracy of decision tree : 0.7

decision tree predict : [0 2 0 2 2 2 1 2 1 0 0 2 2 0 2 0 0 1 2 2 1 1 1 1 0 2 0 0 2 1 1 1 1 1 2 0 0
 2 1 0 0 0 1 2 2 2 0 2 1 0 2 1 1 2 0 2 2 2 1 0]
real : [0 2 0 1 2 2 1 2 1 0 0 2 1 0 2 0 0 1 1 1 1 1 2 2 0 1 0 0 2 2 2 1 1 2 1 0 0
 2 1 0 0 0 0 1 1 1 0 2 1 0 2 2 2 2 0 1 2 2 1 0]
-----
the accuracy of knn : 0.8166666666666667

knn predict : [0 2 0 2 2 2 1 2 1 0 0 2 2 0 2 0 0 1 2 2 1 1 1 1 0 1 0 0 2 2 2 1 1 1 1 0 0
 2 1 0 0 0 0 2 2 1 0 2 1 0 2 1 2 2 0 1 1 2 1 0]
real : [0 2 0 1 2 2 1 2 1 0 0 2 1 0 2 0 0 1 1 1 1 1 2 2 0 1 0 0 2 2 2 1 1 2 1 0 0
 2 1 0 0 0 0 1 1 1 0 2 1 0 2 2 2 2 0 1 2 2 1 0]

```

最後訓練完成後的準確率。由於訓練的資料並沒有很多，因此在準確率表現較差了一點。因為測試的資料也沒有很多，但是如果單純以色塊板來做分析的話，**decision tree** 在分類上(最大深度為 6)感覺較沒有像以人類分類得角度來的直觀，出現了較多的極端值，若測試的資料再多一點的話，這次的訓練資料的預測結果的準確率還會再下降很多，而 **knn**(最近的 15 個鄰居)及 **naïve** 則比較完整地將資料分成了 3 類。

改善方法: 此次只選用了兩個 **features**，若再多選了一些 **features** 在 **decision tree** 的表現應該可以再好一點。