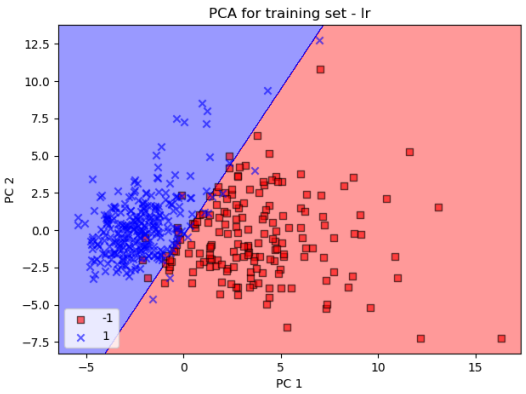
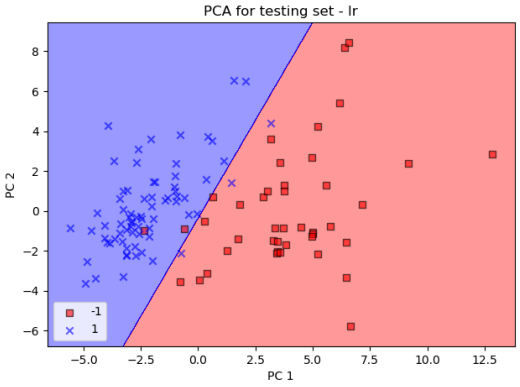
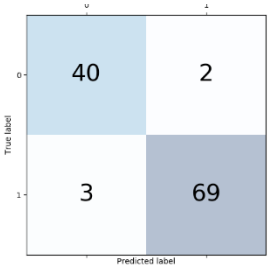
Breast Cancer Classification Using Machine Learning

Ⅰ. Data Modeling with Different Classifier

1. Logistic Regression :

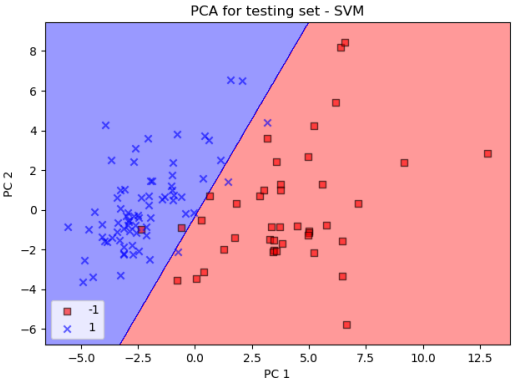


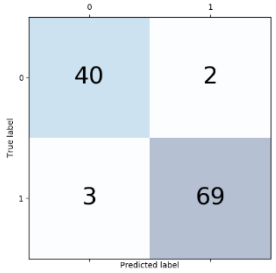




accuracy:0.956

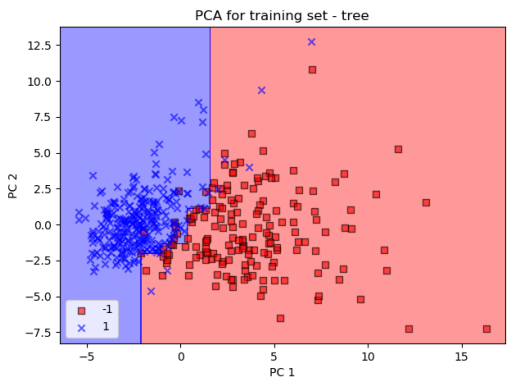
2. SVM :

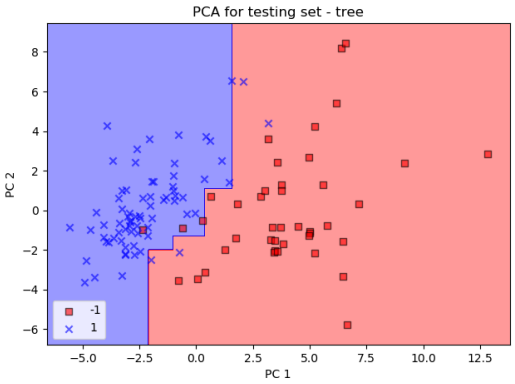


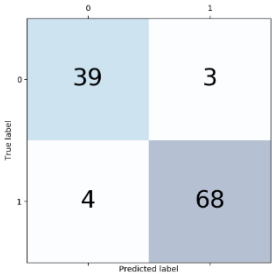


Accuracy : 0.956

3. Decision Tree :

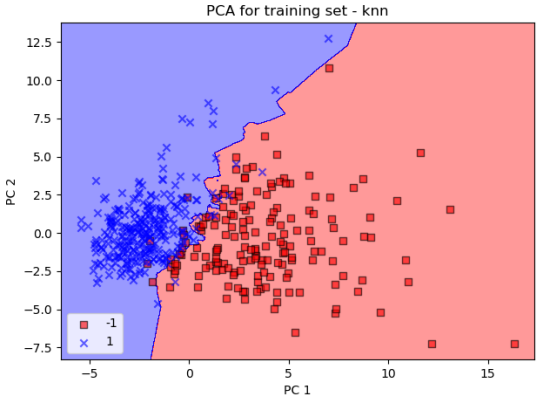


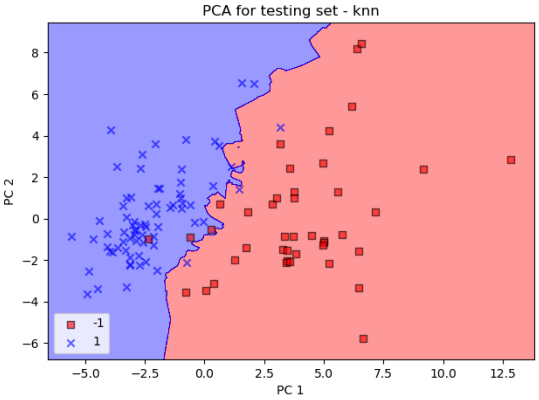


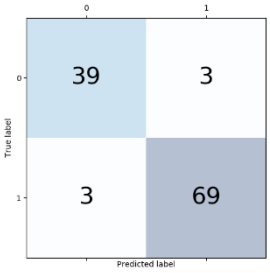


Accuracy : 0.938

4. KNN :







Accuracy : 0.947

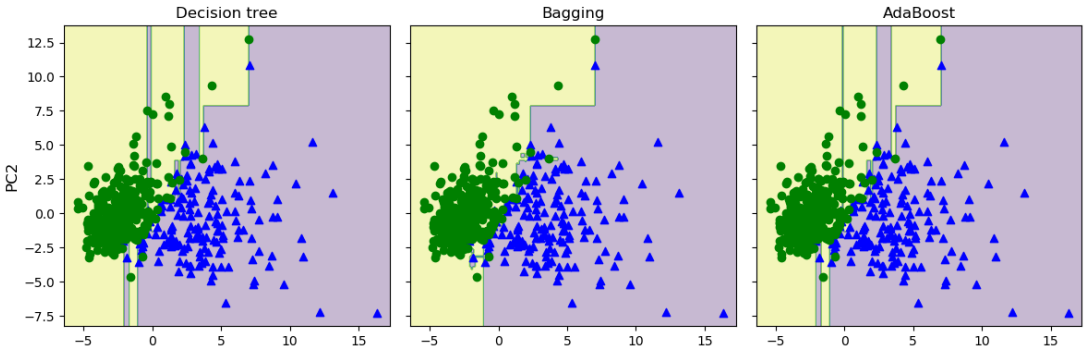
Ⅱ. Compare the Accuracy

Classifier	Logistic Regression	SVM	Decision Tree	KNN
Accuracy	0.956	0.956	0.938	0.947

比較各種分類器的準確度，發現 Logistic Regression 和 SVM 的預測效果較好。

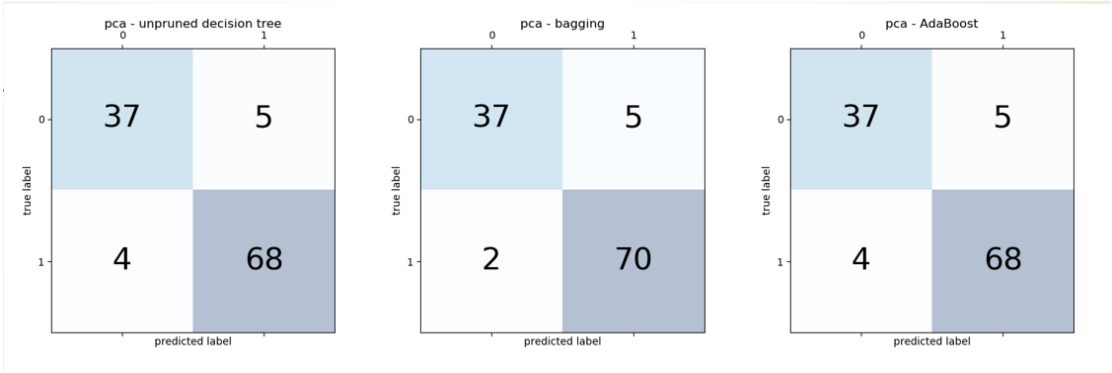
Ⅲ. Combining Different Models for Ensemble Learning

以未限制深度的決策樹來當作 Base estimator



Accuracy = 0.921　　Accuracy = 0.939　　Accuracy = 0.921

- Confusion matrix :



- Compare the Accuracy

Classifier	Accuracy	Classifier	Accuracy
Logistic Regression	0.956	Unpruned Decision tree	0.921
SVM	0.956	Baggings	0.939
Decision Tree	0.938	AdaBoost	0.921
KNN	0.947		

- Conclution:
  1. SVM 和 Logistic Regression 是目前這些分類器當中最有效果的。
  2. Ensembling learning 的方法會增加計算的複雜度和花費更多的成本，但獲得的效能不一定會很顯著，而且如果準確度已經很高，有可能整體方法會不增反降。