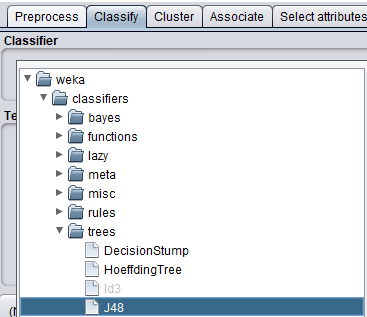
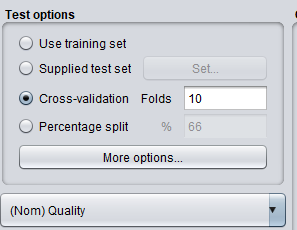
ECT HW8

**Weka Part:**

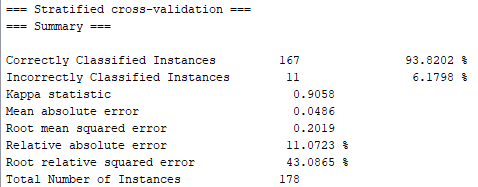
**(a)**

****

* 至Classify中，找到題目要求的J48

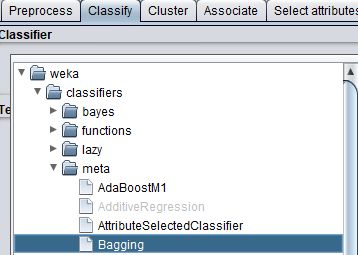


* 依照題目要求Folds = 10，其餘皆使用默認參數

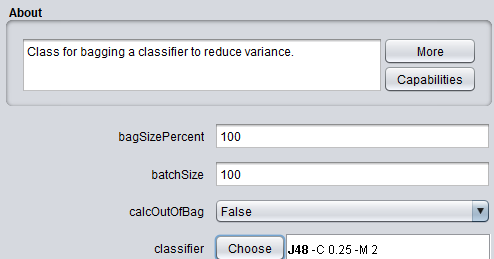


* 分析結果如圖所示，有93.8202%正確率、RMSE = 0.2019

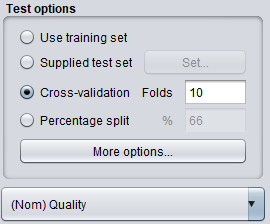
**(b)**



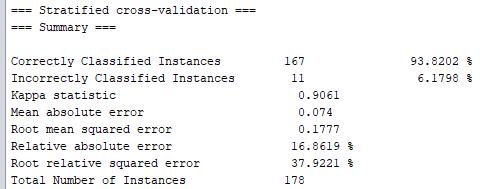
* 在meta底下找到題目要求的Bagging



* 將其classifier設定為題目要求的J48，其餘使用默認選項

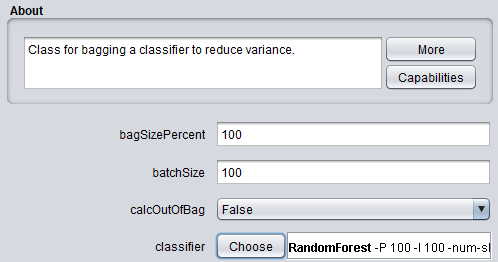


* Folds仍然設為10

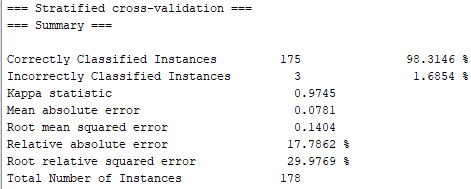


* 雖然準確率跟(a)小題一樣，仍然是93.8202%，但Root Mean Square Error有下降至0.1777，代表他其實是有進步的，雖然不明顯。

**(c)**

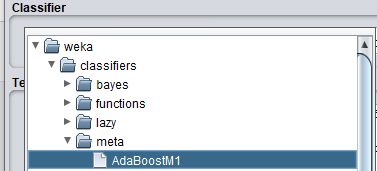


* 依照題目要求，這次一樣用Bagging，但classifier改成RandomForest

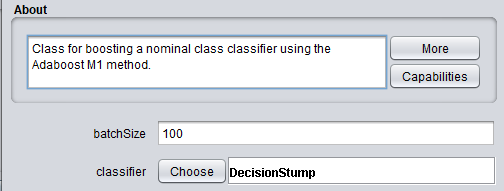


* 可以看出跟前面2個相比，準確率大大的上升了，RMSE也穩定下降中

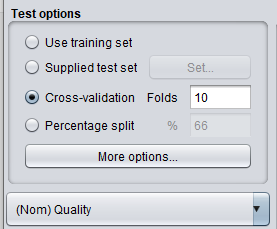
**(d)**



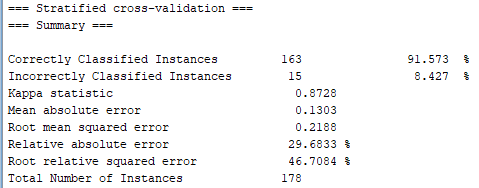
* 一樣繼續在meta底下，找到題目要求的AdaBoost，Weka中叫做AdaBoostM1。



* Classifier選擇題目要求的DecisionStump
* 可以從About中看出，他的確是Adaboost的方法，只是用M1 Method



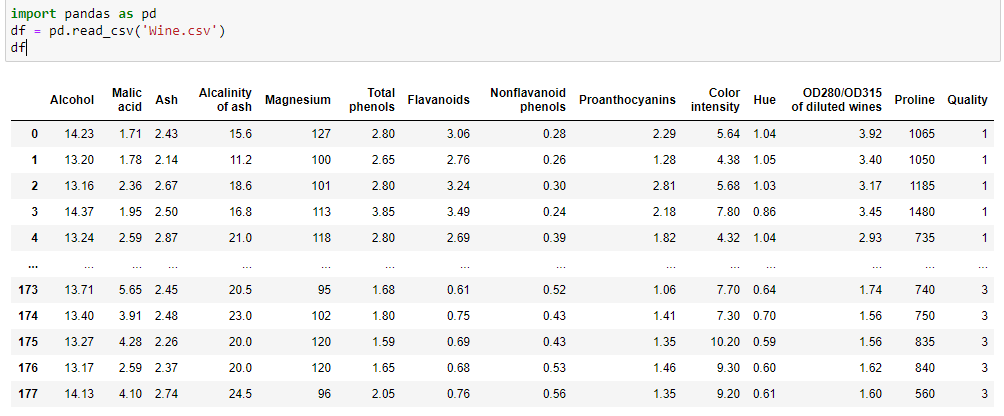
* Folds仍然是10



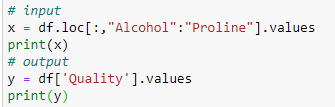
* 準確率比起之前任何方法都低，來到了91.573%，RMSE也提高至0.2188，可見得這次的處理不是個好的選擇

**Python Part:**

**前置處理:**



* 載入資料集

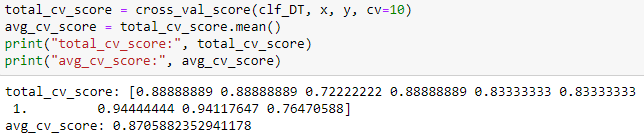


* 將資料集切分為input、output，因題目要求target為Quality，所以output為資料集裡面的「Quality」，其他為input

**(a)**

****

* Import相關套件，並建立DecisionTreeClassifier。
* 在此份作業中random\_state都設為0，因為題目中並未要求seed設為多少，為求方便並告知了解seed的概念，因此不是使用默認，而是統一設為0

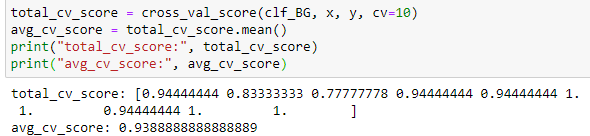


* 使用cross\_val\_score()並設置cv = 10代表10 Folds cross-validation。
* 第1個參數為Classifier，其次為input、output
* 因為用這個方法算出來的是每個Folds的準確率，依照上課所學，會把這些數值平均起來當作準確率，因此調用mean()函數
* 如上圖所示，平均準確率 = 0.8705882352941178

**(b)**



* Import套件，並建立BaggingClassifier
* 依照題目要求n\_estimator = 10, random\_state如(a)小題所提及，設為0
* 因為題目並未要求指定base\_estimator，因此使用默認的decision tree
* 

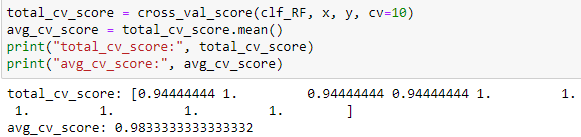


* 一樣使用10 Folds cross-validation，並印出所每個Fold的準確率、平均準確率
* 平均準確率 = 0.9388888888888889，比(a)進步了許多，推測是因為使用Bagging參考多次平均的原因

**(c)**



* Import套件，並建立RandomForestClassifier，random\_state = 0

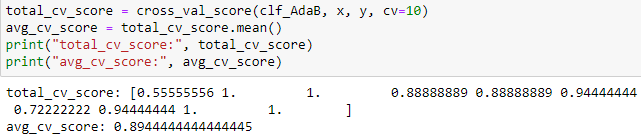


* 一樣進行10 Folds cross-validation分析
* 平均準確率來了0.9833333333333332 !比前面兩提高了許多，已經接近100%了，可見在此例中使用RandomForestClassifier是一個好的方案，可以多方嘗試看有沒有機會至100%

**(d)**



* Import套件，並建立AdaBoostClassifier
* 依照題目要求n\_estimators=10, random\_state仍設為0



* 進行10 Folds cross-validation分析
* 平均準確度只有0.8944444444444445，帶比起(a)小題純粹使用DecisionTreeClassifier得到0.8705882352941178，還是有些許的上升