**Multi-class classification**

1. 使用Tool:weka
2. 前處理:將資料standardize/normalize

原因:減少每個attribute資料分布差異之因素

1. 演算法:
2. J48(C4.5):

工作原理:用屬性計算gain ratio ,並選擇gain ratio 最大的屬性生成node,遞迴進行。

參數＆結果：

|  |  |  |  |
| --- | --- | --- | --- |
| confidence factor | normalize | standardize | score |
| 0.25 | x | x | 0.93939 |
| 0.9 | x | x | 0.93939 |
| 0.9 | o | x | 0.87878 |
| 0.99 | x | o | 0.89393 |
| unprune | x | x | 0.93939 |

1. LMT

工作原理: 用於建立 "邏輯模型樹 "的分類器，為在葉節點上有邏輯回歸函數的分類樹。

參數＆結果：

|  |  |
| --- | --- |
| minNumInstance | score |
| 3 | 0.90909 |

1. Hoeffding

工作原理: Hoeffding tree (VFDT) 是一個增量的、隨時的決策樹歸納算法，能夠從大量的數據流中學習，假設產生例子的分布不隨時間變化

參數:weka預設

結果：0.90909

1. RandomForest

工作原理：結合多顆CART樹（CART樹為使用GINI算法的決策樹），並加入隨機分配的訓練資料，以大幅增進最終的運算結果

參數＆結果：

|  |  |
| --- | --- |
| 100iter | 10000iter |
| 0.93939 | 0.92424 |

1. RandomTree

工作原理：構建一棵樹的類別，考慮在每個節點上隨機選擇K個屬性

參數:weka預設

結果：0.93939

1. REPTree

工作原理：

快速決策樹學習器。使用information gain/variance建立decision/regression tree，並使用reduced-error pruning進行修剪（backfitting）。只對數字屬性的值進行一次排序。處理missing values的方法是將相應的instances分割。

參數：weka預設

結果：0.87878