Machine Learning HW5

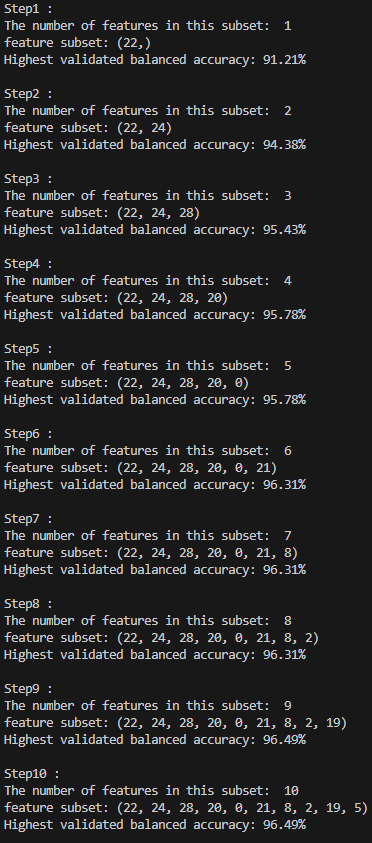
312512005 黃名諄

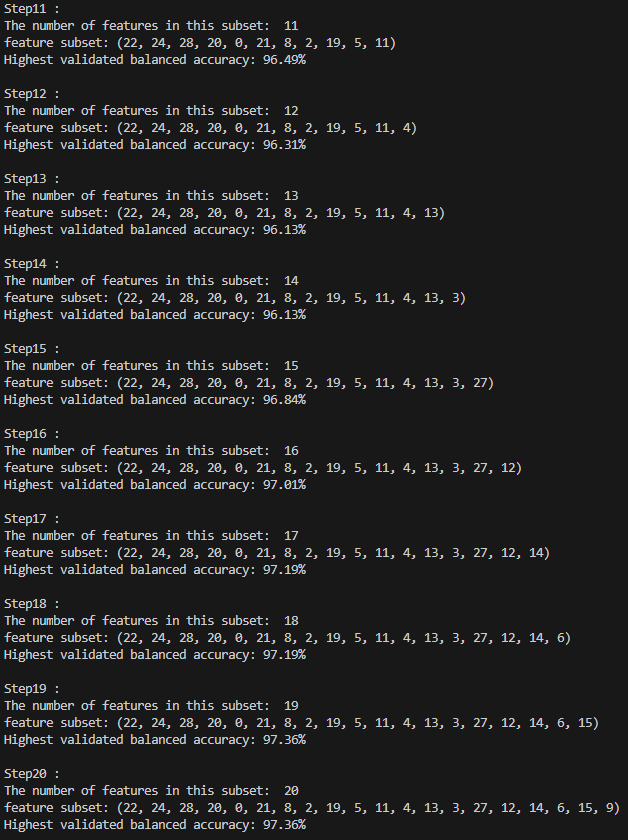
1. Sequential Forward Selection, SFS :

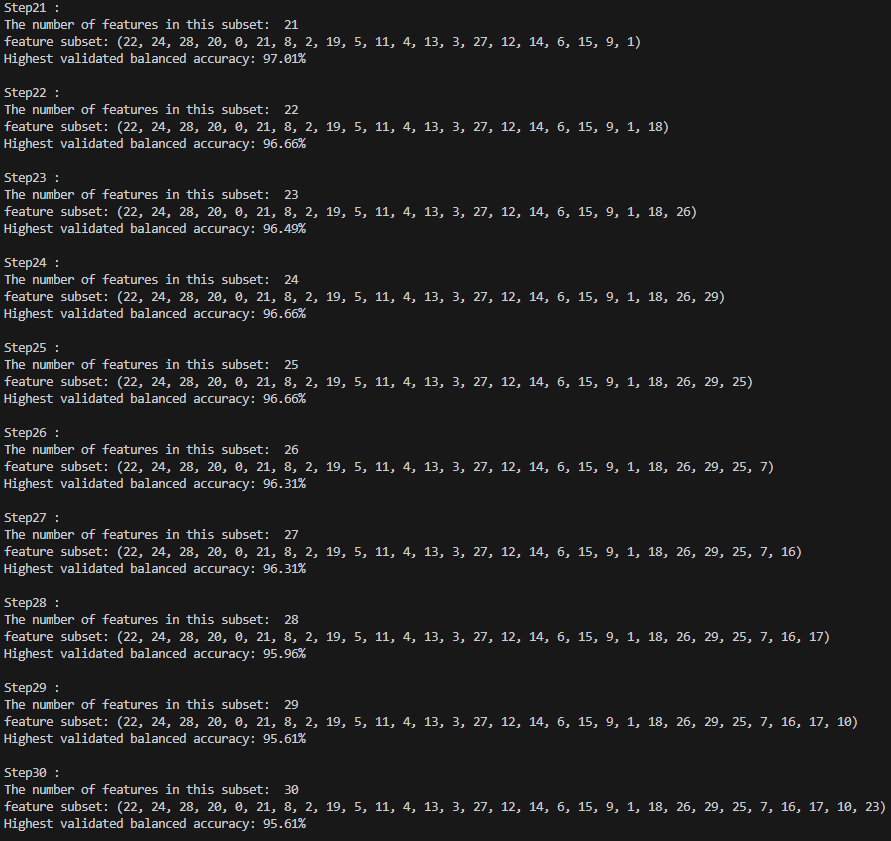
|  |  |
| --- | --- |
| Index | Feature names |
|  | |

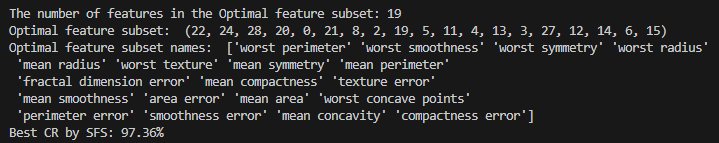
上表是各feature 及其對應index的關係，後面的結果中會使用index代替feature name; 後面結果中分類率都取到小數點後第二位

* SFS Highest validated balanced accuracy result of each step :

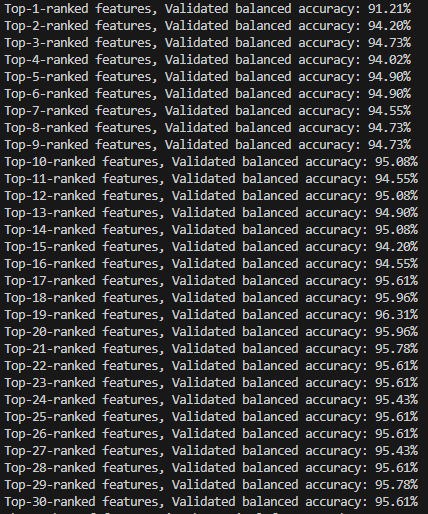
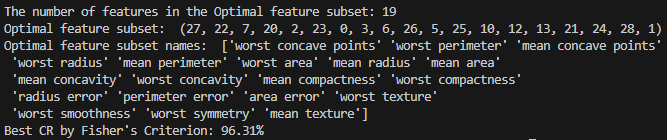






* SFS optimal feature subset :

1. Fisher’s Criterion :

* Top-N-ranked features 之 Validated balanced accuracy :
* Fisher’s Criterion optimal feature subset :

1. 結果討論 :
   * 1. Sequential Forward Selection 和 Fisher’s Criterion 分別屬於 Filter-based 和 Wrapper-based 中的何種特徵篩選方法？Ans :

根據上課所學，Sequential Forward Selection 是Wrapper-based的特徵篩選方法，而Fisher’s Criterion 是Filter-based的特徵篩選方法

* + 1. 一般來說 Filter-based 和 Wrapper-based 各有什麼性質或優缺點？

Ans :

1. Filter-based :

Filter-based方法使用預先設立的指標去評估每個feature的重要性，像Fisher法是計算F-score，而這樣的方法不需要計算分類率，也因此不會受到選擇不同機器學習算法的影響，並且計算簡單而易於實作，但也有缺點，其只考慮各單一特徵的表現，而沒能考慮到特徵之間互相作用的表現，或許兩個weak features一起用卻會有好的表現，這是Filter-based考慮不到的。

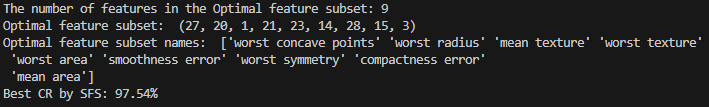
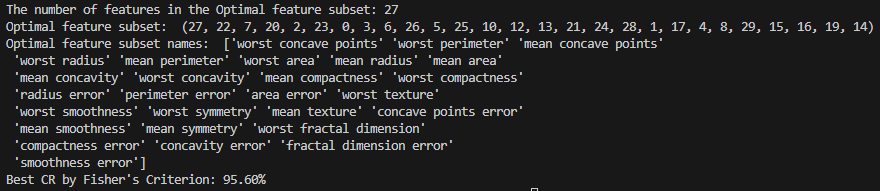
1. Wrapper-based :

Wrapper-based方法實際使用分類率篩選，使用表現直接篩選較為直觀，且其有考慮到不同features組合間的交互作用表現，但也因此綁定了分類器，使用不同分類器可能帶來不同結果，並且計算上不如Filter-based簡單而計算成本較高，另外在過程中被捨棄的features沒辦法再被使用評估，還是可能忽略一些潛在的好特徵組合。

* + 1. 在本次作業的結果中是否有展現出跟上一題你的回答有一致的現象呢? 不管是否一致皆請你試著討論與分析原因

Ans :

在本次作業中能感受到最大差別的是計算效率上，Fisher’s Criterion的計算速度確實快於SFS，因其需要計算2-fold分類率的次數少了很多。另一方面，兩個方法選出的optimal feature subset確實不同，表兩方法對特徵篩選的標準上確實不同。我也發現到更改dataset的2-fold分法，兩個方法的結果都會跟著改變，我認為是因計算分類率上還是會被你的dataset影響到，因為兩者都有計算分類率的環節，都會被不同dataset拆法影響，如下我改了拆分的seed，兩者的結果都和之前不同，但可以發現到SFS選出的最佳特徵集合分類率都較高，可顯示出Wrapper-based如上一題所述考慮特徵間交互作用是很重要的，能選出更佳的特徵組合。

* 更改2-fold分法seed後的SFS結果:
* 更改2-fold分法seed後的Fisher’s Criterion結果: