

機器學習 HW5

系所:電控碩一

學號:313512022

姓名:呂靖樑

Part1

SFS		
Step	feature combination	CR
1	22	92.26
2	22, 24	96.48
3	22, 24, 21	96.66
4	22, 24, 21, 28	97.54
5	22, 24, 21, 28, 0	97.54
6	22, 24, 21, 28, 0, 20	97.71
7	22, 24, 21, 28, 2, 20, 4	97.54
8	22, 24, 21, 28, 2, 20, 4, 10	97.54
9	22, 24, 21, 28, 2, 20, 4, 10, 3	97.54
10	22, 24, 21, 28, 2, 20, 4, 10, 3, 12	97.72
11	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1	97.54
12	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15	97.36
13	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19	97.19
14	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8	97.19
15	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5	97.19
16	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2	97.19
17	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17	97.01
18	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9	96.66
19	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11	96.31
20	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14	96.31
21	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29	96.31
22	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13	96.66
23	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23	96.66
24	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18	96.48
25	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25	96.13
26	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25, 16	95.78
27	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25, 16, 27	95.25
28	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25, 16, 27, 26	95.25
29	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25, 16, 27, 26, 7	95.43
30	22, 24, 21, 28, 2, 20, 4, 10, 3, 12, 1, 15, 19, 8, 5, 2, 17, 9, 11, 14, 29, 13, 23, 18, 25, 16, 27, 26, 7, 6	94.90

Part2

Fisher		
N	Top- N-ranked features	CR
1	27	90.68
2	27, 22	94.38
3	27, 22, 7	92.44
4	27, 22, 7, 20	92.97
5	27, 22, 7, 20, 2	93.85
6	27, 22, 7, 20, 2, 23	94.38
7	27, 22, 7, 20, 2, 23, 0	94.38
8	27, 22, 7, 20, 2, 23, 0, 3	94.73
9	27, 22, 7, 20, 2, 23, 0, 3, 6	94.73
10	27, 22, 7, 20, 2, 23, 0, 3, 6, 26	93.85
11	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5	94.20
12	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25	94.72
13	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10	94.72
14	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12	94.90
15	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13	94.72
16	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21	95.25
17	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24	95.95
18	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28	96.13
19	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1	95.95
20	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17	95.95
21	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4	96.13
22	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8	95.78
23	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29	95.25
24	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15	94.90
25	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16	95.25
26	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16, 19	95.60
27	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16, 19, 14	95.25
28	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16, 19, 14, 9	95.25
29	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16, 19, 14, 9, 11	95.08
30	27, 22, 7, 20, 2, 23, 0, 3, 6, 26, 5, 25, 10, 12, 13, 21, 24, 28, 1, 17, 4, 8, 29, 15, 16, 19, 14, 9, 11, 18	94.90

Part3

1.
SFS 屬於 Wrapper-based 方法，Fisher's Criterion 屬於 Filter-based 方法

2.

方法	優點	缺點
Wrapper-based (SFS)	1. 有考慮特徵交互 2. 能根據模型需求自適應特徵選擇	1. 計算成本較高 2. 不容易通用於不同模型，選擇的特徵對其他模型不一定有效
Filter-based (Fisher's Criterion)	1. 可以用任何算法尋找最佳 N 個特徵 2. 不須用模組來找最佳特徵	1. 特徵之間的相互作用通常不會被考慮 2. 表現可能比 Wrapper-based 差 3. 特徵選擇結果並不一定是針對特定機器學習模型的最佳選擇

3.

比較:

- SFS 的計算成本較高：SFS 需要多次訓練模型，計算成本高共進行了 $(1 + 30) / 2 = 465$ 次 LDA 和 2-Fold CV 計算，而 Fisher's Criterion 僅進行了 30 次。與上題中的優缺點相同。
- 準確率的對比：SFS 的最高分類準確率(97.72%)高於 Fisher's Criterion (96.13%)，這表明 SFS 更能捕捉到特徵之間的交互作用。與上題中的優缺點相同。

原因：

這次作業採用的乳癌數據集具有適中的數據維度以及較強的特徵交互性，因此 SFS 能夠充分發揮其捕捉特徵交互的優勢，表現優於 Fisher's Criterion。然而，若面對資料量更大且數據維度更高的情況，僅使用 SFS 的計算成本可能過高。可以考慮結合兩種方法：先使用 Fisher's Criterion 快速篩選出高分特徵，作為初步特徵子集，再透過 SFS 進一步精細優化，從而平衡效率與性能。

結論:

本次作業結果基本上與前面的分析一致：Wrapper-based 方法在捕捉特徵交互上具有優勢，而 Filter-based 方法更高效但不適合需要深入特徵交互的問題。