1.SFS w/ 2-fold cross validation

Step	Feature No.	Feature Name	Accuracy
1	27	worst concave points	91.04%
2	20	worst radius	94.02%
3	1	mean texture	95.61%
4	21	worst texture	95.61%
5	23	worst area	96.49%
6	14	smoothness error	96.84%
7	28	worst symmetry	97.01%
8	15	compactness error	97.36%
9	3	mean area	97.54%
10	17	concave points error	97.54%
11	5	compactness	97.54%
12	26	worst concavity	97.54%
13	11	texture error	97.54%
14	13	area error	97.36%
15	22	worst perimeter	97.19%
16	4	mean smoothness	97.01%
17	24	worst smoothness	96.66%
18	7	mean concave points	96.84%
19	6	mean concavity	96.48%
20	18	symmetry error	96.31%
21	16	concavity error	96.13%
22	2	mean perimeter	95.78%
23	10	radius error	95.60%
24	8	mean symmetry	95.61%
25	0	mean radius	95.43%
26	9	mean fractal dimension	95.43%
27	19	fractal dimension error	95.43%
28	29	worst fractal dimension	95.43%
29	25	worst compactness	95.43%
30	12	perimeter error	95.08%

Best accuracy is: 97.54%

Features that have best accuracy [27, 20, 1, 21, 23, 14, 28, 15](排序前 8 個)

```
Time cost is 2.6236 sec(s).
```

2. **Fisher' s criterion** w/ 2-fold cross validation

Step	Feature No.	Feature Name	Accuracy
1	19	fractal dimension error	61.68%
2	14	smoothness error	61.51%
3	17	concave points error	73.11%
4	9	mean fractal dimension	72.23%
5	4	mean smoothness	81.19%
6	7	mean concave points	90.16%
7	18	symmetry error	89.63%
8	25	worst compactness	92.62%
9	26	worst concavity	92.27%
10	28	worst symmetry	94.03%
11	20	worst radius	94.03%
12	12	perimeter error	93.85%
13	21	worst texture	95.78%
14	8	mean symmetry	95.08%
15	13	area error	95.43%
16	3	mean area	95.43%
17	23	worst area	95.78%
18	22	worst perimeter	95.43%
19	2	mean perimeter	95.08%
20	1	mean texture	95.25%
21	0	radius	95.25%
22	11	texture error	95.25%
23	6	mean concavity	95.08%
24	10	radius error	95.08%
25	5	mean compactness	94.90%
26	27	worst concave points	94.90%
27	24	worst smoothness	94.73%
28	16	concavity error	94.90%
29	15	compactness error	94.90%
30	29	worst fractal dimension	95.08%

```
Best accuracy is: 95.78%
Selected Features that have best accuracy: [19 14 17  9  4  7 18 25 26
28 20 12 21](Feature indices)
Time cost is: 0.3835 sec(s).
```

問題與討論：

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. Sequential Forward Selection屬於Wrapper methods(包裝器法)，使用預測性的機器學習演算法來評估最佳的特徵子集合性能，進而選出最佳的特徵子集合；Fisher’s Criterion則是Filter-based的特徵篩選方式，不考慮將來使用何種模型進行學習，只考慮變數和預測值的相關性，僅以數據本身特性決定，不考慮聯立機率。

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

Ans. Filter-based的優點是計算效率高，不須使用過度複雜的模型，且對高維的數據通常比現較佳，缺點為忽略特徵之間可能存在的相互作用(聯立機率)，且不考慮不同機器學習演算法之間可能存在的差異，可能篩選出對分類並無顯著幫助的特徵；Wrapped-based的優點為考慮特徵之間的交互關係，以此提高不同機器學習演算法的個別性能，缺點是計算效率較低，需要訓練多個不同模型來評估每個特徵子集合性能，在處理高為度數據的過程中容易產生Overfitting現象。

3. 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.是。

SFS 在前幾步就有著 90%以上的分類正確率，而 Fisher’s criterion 直到 Step 6 才達到 90%的正確率，顯示 Wrapped-based 的方式的確考慮到演算法的不同、以及各數據間的交互影響，較為周全。計算複雜度部分，由執行時間比較，明顯 Fisher’s criterion 運算量較小(2.6 秒 VS 不到 0.4 秒)，同樣符合上一題所說，Wrapped-based 計算效率較差。結論：從正確率和耗時來看，本作業的結果符合預期，不過因為現代電腦運算能力較佳，本次處理數據並不到太大，差距僅有數秒，但仍然看得出些許差異。