- In filter method we only see one column at a time but we don't see two column together.

- Wrapper Methods -
  ↳ $f_1, f_2 \ldots f_n$ y
  We make subsets of features like
  $f_1 f_2$, $f_1 f_2 f_3$ etc
  and on each feature we apply ML algo
  and then calculate r2-score. The subset
  having highest r2 score is selected
  we wrap model based evaluation around feature
  selection

- Steps-
  - Subset Generation
  - Subset Evaluation (r2 score)
  - Stopping criteria

Types -
ↆ Exhaustive feature select^n / Best Subset Sol^n
ↆ forward select^n
ↆ Backward eliminate^n
ↆ Recursive feature eliminate^n

1) Exhaustive feature select^n -
ↆ we make all possible subsets and calculate r2-score of each subset and the one having greatest will be selected.

$f_1 \{_2 y =\}$ $f_1 y$ $f_2 y$ $f_1 f_2 y$
            0.36   0.58   (0.92) ✓

$$\therefore \boxed{2^n - 1}$$

. Disadvantages -
. Computational complexity
. Risk of overfitting
. Requires a good evaluation method
(r2 score works good when we have less score but as we increase no. of columns performance become ~~bad~~ bad)

ↆ This method gives best result when we have less no. of columns (20)

2) Backward Elimination or seq. Back. select^n
ↆ in this we eliminate the feature one by one.

$\boxed{f_1}$ $\boxed{f_2}$ $\boxed{f_3}$ $\boxed{f_4}$ $\rightarrow$ acc = 0.33 (we take all feature)

Now we will remove 1 feature

f_1 $\boxed{f_2}$ $\boxed{f_3}$ $\boxed{f_4}$ $\rightarrow$ 0.61

$\boxed{f_1}$ f_2 $\boxed{f_3}$ $\boxed{f_4}$ $\rightarrow$ 0.71

✓ $\boxed{f_1}$ $\boxed{f_2}$ f_3 $\boxed{f_4}$ $\rightarrow$ 0.91

$\boxed{f_1}$ $\boxed{f_2}$ $\boxed{f_3}$ f_4 $\rightarrow$ 0.65

Now remove are more feature

f_1 $\boxed{f_2}$ $\boxed{f_4}$ $-$ 0.79

$\boxed{f_1}$ f_2 $\boxed{f_3}$ $-$ 0.81

✓ $\boxed{f_1}$ $\boxed{f_2}$ $\boxed{f_4}$ -0.63

Now remove are more feature

$\boxed{f_1}$ f_2 $-$ 0.63

✓ f_4 $\boxed{f_2}$ $-$ 0.63

NN f_1 f_2 f_3 f_4 $\rightarrow$ 0.83

f_1 f_2 f_4 $-$ 0.91 ✓ (selected)

f_1 f_2 $-$ 0.63

f_2 $-$ 0.63

Total $\Rightarrow$ $\dfrac{n(n+1)}{2}$ models

- **Disadvantage –**
  → At every iteration we are doing local selection
    i.e we are removing that feature which
    is not good at that time.
    Suppose we remove f_3 in 1st iteration
    and now we will never use it. What if
    in other iteration $f_3 - f_4$ was best comb.
    As we already removed f_3 we will never
    get best combo

3) Sequential forward Selection –

| f_1 | → 0.63 ✓ | ⟹ | F_1 f_2 → 0.63 |

| f_2 | – 0.51 | | F_1 f_3 → 0.71 |

| f_3 | – 0.43 | | F_1 f_4 → 0.80 ✓ ⟹ f_1 f_4 f_2 – 0.81 |
| | | | f_1 f_4 f_3 – 0.85 ✓ |

| f_4 | → 0.49 | | ↓ |

$$f_1 \, f_4 \, f_3 \, f_2 → 0.83$$

So f_1 → 0.63
$F_1 \, F_4$ → 0.80
$F_1 \, f_4 \, f_3$ → 0.85 ✓ Ans
$F_1 \, f_4 \, f_3 \, f_2$ → 0.83

Exhaustive → $2^n - 1$ model
Backward ⟹ $\dfrac{n(n-1)}{2}$
forward

from 5b3 we want 55D   → Backward
                  5D   → forward

we just want best combn" →
                   ↳ any forward or
                      backward

: Advantages of wrapper method

1) Accuracy → better than filter
2) It consider interactions of features
    (filter method don't)

· Disadvantages -

1) computational complexity
2) Risk of overfitting
3) Model specific