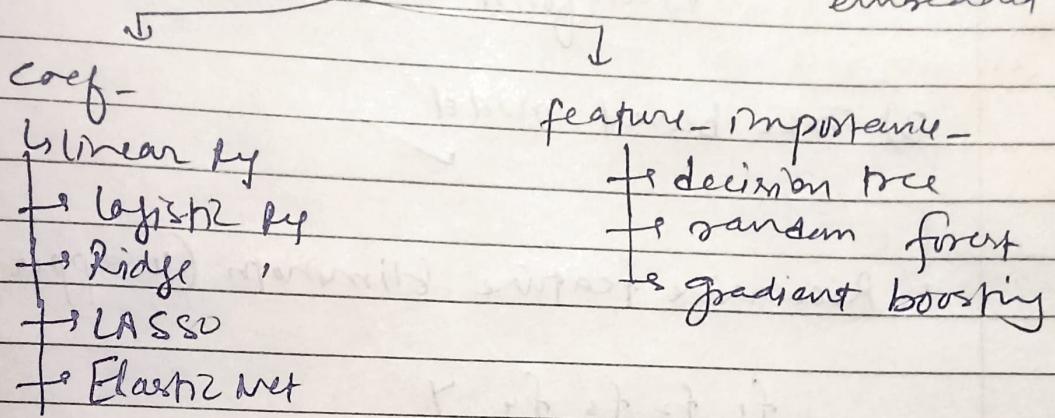


Embedded based feature Selection -

- filter method doesn't do feature selection and wrapper method is computationally slow So to solve these problems we have embedded method.
- In this predicting model is made and at the same time feature importance is also calculated
- Any ML algo which has following can use embedded



i) Linear Reg -

$$lpa = \beta_0 + \beta_1 cgpa + \beta_2 iq$$

β_1 tells if we ↑ cgpa by 1 then lpa how much lpa increase

So, β_1 & β_2 can be treated as coeff - feature importance
coeff - is

But cond'n is β_1, \dots, β_n

- Linearity
- Independence
- Homoscedasticity
- Normality
- No multicollinearity

2) Regularized model

is a linear model that includes penalty term in loss fn.

[] []

Ridge

LASSO

Elastic Net

\Rightarrow Remember $\underbrace{\text{fit}_1, \text{fit}_2}_{\downarrow}$ - transform

in training data

transform \rightarrow on both train & test

3) Tree based model

• Recursive feature Elimination (wrapper based)

$f_1 f_2 f_3 f_4 Y$

↳ choose a model having coeff-/feature-importance

Remove feature having least feature-importance
Let's remove f_4 .

Repeat Same step

$f_1 f_2 f_3$

$f_1 f_3$

$f_1 f_3 \Delta y$
best feature

. Advantages of Embedded-

- Good performance
- Efficient
- less prone to overfitting

. Disadvantage -

- Model Specific
- complexity
- Tuning required (what value of α)
- Stability

. Mutual Information (filter based)

↳ measure of dependency b/w two variable

$$MI = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left[\frac{p(x, y)}{p(x)p(y)} \right]$$

$p(x, y) \rightarrow$ Joint Prob.

$p(x), p(y) \rightarrow$ marginal Prob.

→ MI -

→ is non negative

→ is symmetric $MI(x, y) = MI(y, x)$

→ can capture any kind of statistical dependency