

## Embedded based feature Selectn.

→ filter method doesn't do feature interaction 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

coef-

- + Linear Reg
- + Logistic Reg
- + Ridge
- + LASSO
- + Elastic Net

feature-importance-

- + decision tree
- + random forest
- + gradient boosting

1) Linear Reg.

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

$\beta_1$  tells if we ↑ c\_gpa by 1 then lpa how much lpa increase

So,  $\beta_1$  &  $\beta_2$  can be treated as coef-  
coef- in feature importance

But cond<sup>n</sup> is  $\beta_1, \dots$  etc

→ Linearity

→ Independence

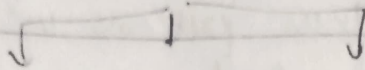
→ Homoscedasticity

→ Normality

→ No multicollinearity

## 2) Regularized model

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



Ridge

LASSO

ElasticNet

⇒ Remember  $\underbrace{\text{fit} \cdot \text{fit} \cdot \text{transform}}_{\text{in training data}}$

transform → 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 steps

$f_1 \ f_2 \ f_3$

$f_1 \ f_3$

$f_3$  is  
best feature



## Advantages of Embedded-

- good performance
- Efficient
- less prone to overfitting

## Disadvantage-

- model Specific
- complexity
- Tuning Required (what value of  $\alpha$ )
- stability

## Mutual Information (filter based)

↳ measure of dependency b/w two variable

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

$P(x, y)$  → Joint Prob.

$P(x), P(y)$  → marginal Prob.

→ MI -

→ is non negative

→ is symmetric

→ can capture any kind of statistical dependency

$$MI(x, y) = MI(y, x)$$