Page No.
Date

- Embedded based feature Selectn.

→ filter method doesn't do feature interactn 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.

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

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

So, β_1 & β_2 can be treated as coef_ feature importance

coef_ is

But condtn is β_1 .. etc

- → Linearity
- → Independence
- → Homoscedasticity
- → Normality
- → No multicollinearity

2) Regularized model
is a linear model that includes penalty term in loss fn

Ridge   LASSO   Elastic Net

=) Remember fit , fit_transform
$\underbrace{\text{fit , fit\_transform}}_{\text{on training data}}$

transform → on both train & test

3) Tree based model ✓

- Recursive feature Elimination (wrapper based)

f1 f2 f3 f4 Y
↳ choose a model having coef_ / feature_importance_

Remove feature having least feature_importance_
Let's remove f4

Repeat same steps
f1 f2 f3

f1 f3

f3 → Y
↳ best feature

Page No.
Date

- 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

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

$P(x,y)$ → Joint Prob.
$P(x), P(y)$ → marginal Prob.

- → MI -
  - → is non negative
  - → is symmetric $MI(X,Y) = MI(Y,X)$
  - → can capture any kind of statistical dependency