

FEATURE SELECTION

→ Backward Elimination Technique using OLS

→ Feature Selection Techniques

Linear regression (50_startups.csv)
R&D, Adm, market, state, Profit
features categorical label

California, Florida, NY, R&D, Adm, market → features

$$\text{Profit} = b_0(1) + b_1(\text{California}) + b_2(\text{Florida}) + b_3(\text{NY}) + b_4(\text{R\&D}) \\ + b_5(\text{Adm}) + b_6(\text{market})$$

Option → To select the best features using multiple approaches.

- ① Backward Elimination Technique using OLS (Ordinary Least Squares)
- ② Perform Recursive Feature Elimination (RFE)
- ③ Get Best features Based on model (algo)
- ④ Get Best features using ANOVA (Analysis of Variance)

① Backward Feature Elimination using OLS.

OLS \rightarrow Ordinary Least Squares \rightarrow deploy model not possible
Linear Regression \rightarrow deploy model is possible

Steps (Algo):

- ① Perform All-IN (input all features and get the summary)
- ② Decide the Significance Level (S.L.) \rightarrow 0.01
0.05
0.1
 \downarrow
accepted error %age \rightarrow 5% acceptable error
95% confidence in my model
- ③ Check which feature in summary has highest p-value.
- ④ If p value $>$ S.L., eliminate that feature. else go to ⑤
- ⑤ Repeat from ②.
- ⑥ consider all features has importance for model creation