Embedded Methods

Embedded methods are feature selection technique which perform feature selection as part of the model construction process. They are called embedded methods because feature selection is embedded methods because feature selection is embedded within the construction of the machine learning model. These methods aim to solve the limitation of filter and wrapper methods by including the interaction of the feature while also being more computationally efficient.

Problem

filter ↓ feature interaction individual

wrapper ↓ slow

f1 f2

feature interaction individual $\xrightarrow{\text{Solution}}$ embedded

slow $\xrightarrow{\text{Solution}}$ embedded

embedded ↓ feature importance

model → data → feature importance

model → prediction

model → feature importance

machine learning

coef_
- Linear Regression
- Logistic
- Ridge
- Lasso
- Elastic Net

feature Importance
- decision tree
- random forest
- gradient Boosting

Linear Regression

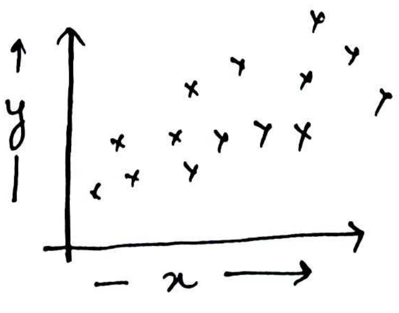
gpa | iq | lpa

$$lpa = \beta_0 + \boxed{\beta_1}\, gpa + \boxed{\beta_2}\, iq$$

Importance important

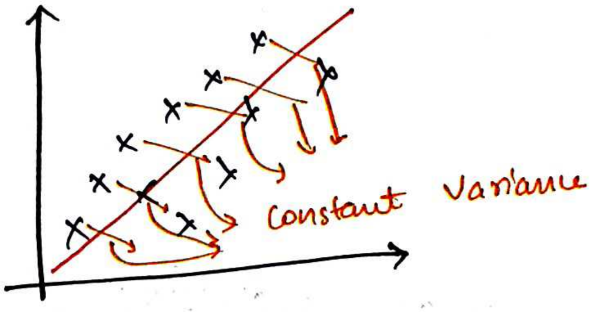
Assumption:

1. Linearity :- The relationship between the independent and dependent variables is linear. This is also means the change in the dependent variable for a unit change in the dependent variable(s) is constant.

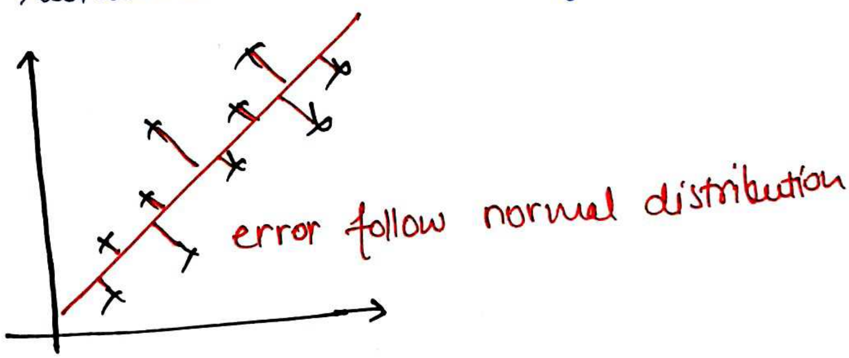


2. Independence : The observation are independent of each other. The implies that the residuals (the difference between the observed and predicted values) are independent.

3. Homoscedasticity: The variance of the residuals is constant across all levels of the independent variables.



4. Normality : The residuals are normally distributed.



5. No Multicollinearity :- The independent variables are not highly correlated with each other. This assumption is really important when you want to interpret the regression coefficients.

Regularized Models

Regularized linear models are linear models that include a penalty term in the loss function during training. The penalty term discourage the learning of a too complex model, which can help prevent overfitting.

- Ridge
- Lasso → feature selection
- Elastic NET

Recursive Feature Elimination

wrapper → hybrid

f_1 f_2 f_3 f_4 Y → model → coef / feature important

f_4 → low

best Rfe

f_1 f_2 f_3 Y → model → feature

f_2 → low

f_1 f_3 Y → model → feature

f_1 → low

$\boxed{f_3 \rightarrow \text{best feature}}$

Advantage and Disadvantage

Advantage:

1. **Performance**: They are generally more accurate than filter methods since they take the interaction between features into account.

2. **Efficiency**:- They are more computationally efficient than wrapper methods since they fit the model only once.

3. **Less Prone to Overfitting**: They introduce some form of regularization, which help to avoid overfitting. For example, Lasso and Ridge regression add a penalty to the loss function, shrinking some cofficient to zero.

Disadvantage

1. **Model Specific**:- Since they are tied to a specific machine learning model, the selected features are not neccessarily optimal for other models.

2. **Complexity** :- They can be more complex and harder to interpret than filter methods. For example, understanding why Lasso shrinks some coefficients to zero and not others can be non-trivial.

3. **Tuning Required** : They often have hyperparameter that need to be tuned, like the regularization strength in Lasso and Ridge regression.

4. **Stability** : Depending on the model and the data, small changes in the data can results in different sets of selected features. This is especially true of models that can fit complex decision boundaries like descion trees.