## Feature Selection Techniques

- Feature selection and cleaning dataset is the first aim before the machine learning.
- Because if we go for the machine larning with the some irrelevant feature that time we will get the undesirable OUTPUT or negetive we can say.
- So to get the more precise output or model we have to remove the some unwanted feature from datasets which are causing the Negetive impact on the model.
- If we work for the unwated data and train model for them it leads to waste of your time.

| Filter methods | Wrapper methods | Embedded methods |
|---|---|---|
| Generic set of methods which do not incorporate a specific machine learning algorithm. | Evaluates on a specific machine learning algorithm to find optimal features. | Embeds (fix) features during model building process. Feature selection is done by observing each iteration of model training phase. |
| Much faster compared to Wrapper methods in terms of time complexity | High computation time for a dataset with many features | Sits between Filter methods and Wrapper methods in terms of time complexity |
| Less prone to over-fitting | High chances of over-fitting because it involves training of machine learning models with different combination of features | Generally used to reduce over-fitting by penalizing the coefficients of a model being too large. |
| Examples – Correlation, Chi-Square test, ANOVA, Information gain etc. | Examples - Forward Selection, Backward elimination, Stepwise selection etc. | Examples - LASSO, Elastic Net, Ridge Regression etc. |

## Why Feature Selection ?

- Shorter the training time
- Eaiser to interpret model - by the diamesionality from high to low.
- Reduce Overfitting
- Improve the Accuracy

# Feature Selection Method :-

- Filter Methods.
- Wrapper Methods.
- Embedded Methods.

## Filter Methods.

- Its done by using the preprocessing method.
- The selection of feature is indepnded on any machine learning algorithm in this filtering method.
- Feature are the selected on the basis of their score ,various stastical test,co-relation with the outcome variable.
- Those stastical test can be done by with the **chi-square,fisher square,anova,mutual information,variance.**
- It uses the individual feature predictive power.
- It is less computationally expensive.
- Uses the corelaton power on feature to target space.
- Usefull to quick screening.

## Wrapper Methods.
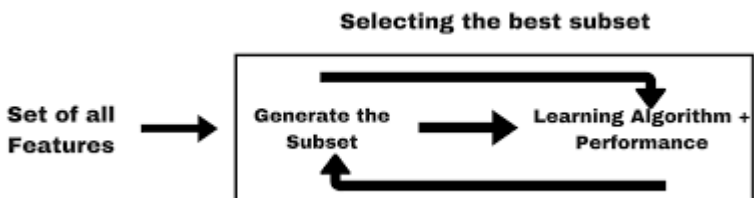


- In the warpper method we try to use the subset of the features and train our model by using those.
- Based on the inferrence that we draw from the previous model we decide to add or remove features from the subsets.
- The problem essentially reduce to search problem actually that means.
- There is lots of the feature in that we need to select subset of the feature where we can maximiz the preformace of our model.
- Since,we trying to do with the help of the combination of the avaliable features.
- Use the combinations varibles to determine predictive power.
- Find the best combination of the variable.
- Its computationally expensive than the filter method.
- Its not recommnded on the High numbers of features.

## Embedded Methods.

- It combining the qualities of the filter and wrapper method.It implementing the with the algorithms that have their own building.\
- Feature selection methods some of the mose popular example:- LASSO,RIDGE regression and Tree importance.
- **Lasso regression -** It perforce the l1 regularization which add the penalty equivalent to the absolute value of the magnitude of the coefficient.
- **Ridge regression:-** It performs the L2 regularization which adds the penalty to the equivalent to the square of the magnitude of the coefficient.
- **Tree Importance:-** It gives the score of each features and uses tree based classifier to gets the importance of each individual features.
- Embedded method is faster than the wrapper method
- It is more accurate than the filter method.
- It performce the feature selection during the algorithm implementation that why it known as 'embedded'.

## Steps For The Wrapper Method.

- Cycle of the wrapper is method is below way..
  - Set of the all features which are avalible in the datasets.
  - Generate the subsets.
  - Train the subsets to the ml algorithms.
  - After we will get the performance.

## Types Of The Wrapping Methods.

- Subset Selection(Exhaustive feature selection)
- Foreward Step Selection
- Backward step selection (Recursive Feature selection)

### Foreward Step Selection.

- It creates the number of the subset of the feature.
- Machine learning algorithm try to compute the performance of evey feature subset individualy and then choose one of them whoever give us the best performance.
- Afterwards,it tries to get the combinations with the remaing features (like the feature which did not give us the better preformace last time)then one of whihch combination whoever will give us te best performance we have to select it and do it same way untill and unless we will not get the desirable feature where we want to put is
- Overall,It steping the foreward manner and turn by turn use the ml algorithm to train the model to get the best feature set.
- **Example**
- If we having the 100 features and if we want to select only top 10 feature so in that we will go for this one from 1 to 10 step by step.

### Backward step selection.

- First we try to get the performance of all the features by using the ml model and give them trainng and then we try to remove.
- One featute then we check the performce if not get the required then we have to same thing again and again until and unless we will get the desired features selection.
- **For an Example :** we having the 100 features and from which we select the only ten features.
- So for that it will get the start in this case from the 100 and then step by step reducing the feature 99,98... upto the 10 features.
- We will get the 10 features which will give us the maximum .

### Wrapper Methods(Exhaustive feature selection)

- Wrapper method is subset selection method.
- Fit the model with each possible combinations of N features.
- It's required the massive computational power generally its avoided in the machine algorithms because it not in use right now.

# Difference Between The Filter And Wrapper Method.

- **Filter method :-** It measure the relevance of the feature by their **co-relation with the depended variable that means featured space.**
- While **wrapper method** is use to measure usefullness of subset of feature **by actually training a model on it.**
- That means in the **filter method doesnt apply any machine learning algorithm.**
- But the **wrapper uses machine learning algorithm to test usefullness of the of the selected feature.**
- Fiter method is much faster compare to the wrapper method as they do not use ml algorithm and they do not try to train any model.
- On the other hand the **wrapper computationally very expensive and try to do train the model everytime.**
- Fiter method use the **some stastical methods to features while wrapper uses the cross validation.**
- Filter method might failed to find the bset subset of the feature many occasion.
- But wrapper method can always provide best subset of the feature and using the subset of the feature of the wrapper method.
- It makes the model more prone to the overfitting as compare to the using the subset of the feature with the help of the filter method.