Feature Selection

### Feature selection is a process that chooses a subset of features from the original features so that the feature space is optimally reduced according to a certain criterion.

#### There are three general classes of feature selection algorithms: Filter methods, wrapper methods and embedded methods.

# Role

1. To reduce the dimensionality of feature space.

2. To speed up a learning algorithm.

3. To improve the predictive accuracy of a classification algorithm.

4. To improve the comprehensibility of the learning results.

# Methods

1. **Filter methods**

Used while doing the pre-processing step. These methods select features from the dataset irrespective of the use of any machine learning algorithm.

Used techniques: Information Gain, Chi-square test, Correlation Coefficient.

### Wrapper methods

Train the algorithm by using a subset of features in an iterative manner. Based on the conclusions made from training prior to the model, addition and removal of features takes place.

Used techniques: Forward selection, Backward Elimination, Bi-directional elimination.

1. **Embedded methods**

The feature selection algorithm is blended as part of the learning algorithm, thus having its own built-in feature selection methods. Embedded methods encounter the drawbacks of filter and wrapper methods and merge their advantages.

Used techniques: Regularization, Tree-based methods.