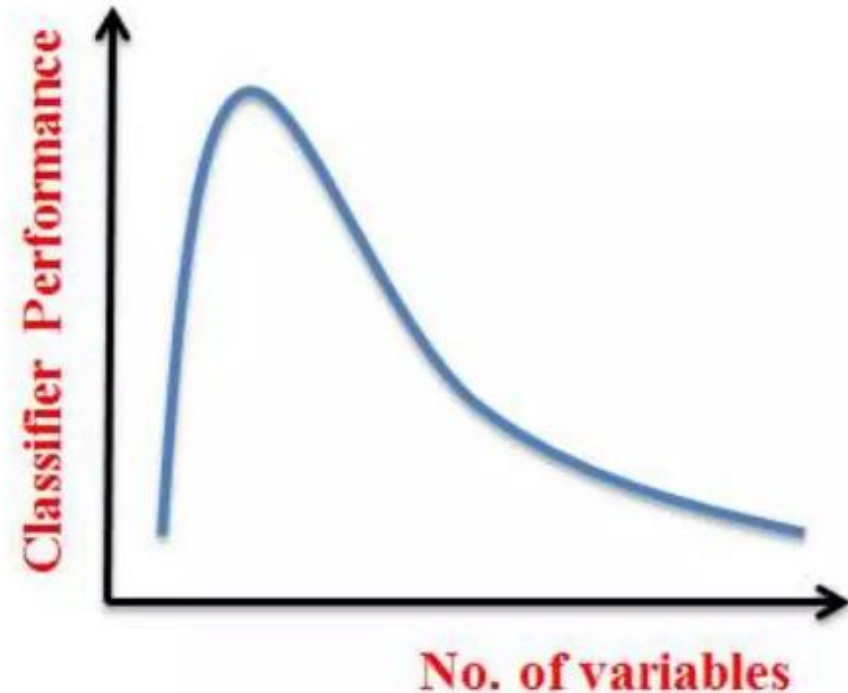# Feature Selection

Feature Selection

- It is the process of selecting a subset of relevant features for use in model construction.

- It is also called variable selection or attribute selection.

- It reduces the complexity of model.

- It either improve or maintain accuracy of model.

Curse of Dimensionality

- The amount of data required to achieve the same level of accuracy increases exponentially as the number of features increases.

- But in practice, the volume of training data available to us is fixed. Therefore, in most cases, the performance of the classifier will decrease with an increased number of variables.
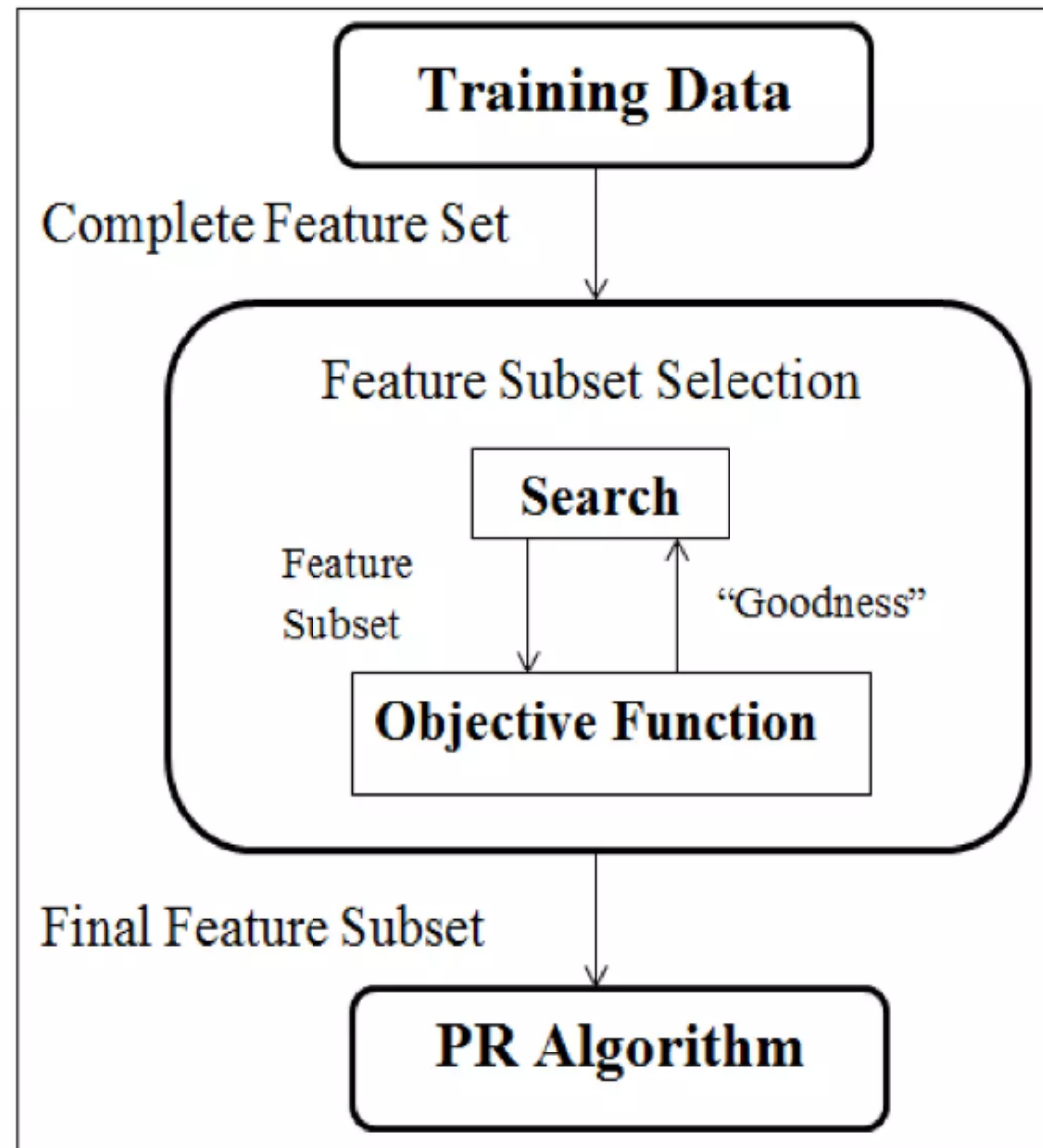
Feature Selection Steps

Feature Selection is an optimization problem.

Step 1 : Search the space of possible feature set.

Step 2 : Pick the subset that is optimal or near optimal with respect to some objective function.

Feature Selection Methods

- Feature Selection Algorithms can be broadly classified into 3 classes :
  - Filter Methods
  - Wrapper Methods
  - Embedded Methods

Filter Methods

- purely Statistical in nature.

- used as a preprocessing step.

- the selection of features is independent of any machine learning algorithms.

- features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable.

		Feature	
		Categorical	**Continuous**
Response	**Categorical**	Chi-Square	Anova
	Continuous	LDA	Pearson Correlation

- **LDA:** Linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes (or levels) of a categorical variable.

- **ANOVA:** ANOVA stands for Analysis of variance. It is similar to LDA except for the fact that it is operated using one or more categorical independent features and one continuous dependent feature. It provides a statistical test of whether the means of several groups are equal or not.

- **Chi-Square:** It is a dependency test which helps us to determine either there is any kind of dependence among the variables under observation or not.

- **Pearson's Correlation:** It is used as a measure for quantifying linear dependence between two continuous variables X and Y. Its value varies from -1 to +1.

Wrapper Method

- work on the basic principles of Combinatorics.

- the learning algorithm (classifier) itself is used to perform the Feature Selection and select the top features.

- Examples :

  - forward feature selection

  - backward feature elimination

  - recursive feature elimination

- **Forward Selection:** Forward selection is an iterative method in which we start with having no feature in the model. In each iteration, we keep adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model.

- **Backward Elimination:** In backward elimination, we start with all the features and removes the least significant feature at each iteration which improves the performance of the model. We repeat this until no improvement is observed on removal of features.

- **Recursive Feature elimination:** It is a greedy optimization algorithm which aims to find the best performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination.

Embedded Methods

- combine the qualities of filter and wrapper methods.

- dynamic Feature Selection Methods.

- Based on the inferences that we draw from the previous model, we decide to add or remove features from your subset. The problem is essentially reduced to a search problem.

Feature Selection vs. Extraction

dimensionality reduction

- Selection: choose a best subset of size m from the available d features

- Extraction: given d features (set *Y*), extract m new features (set *X*) by linear or non-linear combination of all the d features

- New features by extraction may not have physical interpretation/meaning

- Examples of linear feature extraction
  - Unsupervised: PCA; Supervised: LDA

- Criteria for selection/extraction: either improve or maintain the classification accuracy, simplify classifier complexity