WELCOME TO

TECH I.S.

Feature Selection

Feature Selection

In many applications, we often encounter a very large number of potential features that can be used.

Which subset of features should be used for the best results?

Need for a small number of features:

To avoid "Curse of dimensionality" : Data needed to train classifier grows exponentially with no. of dimensions leading to Overfitting and Generalization performance

To avoid "Reduced complexity and run-time" : Time taken for building a model increases as no.of features increase.

Feature Selection vs. Feature Extraction

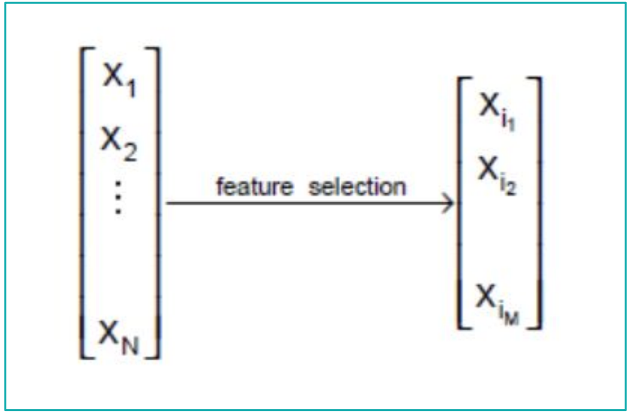
Two general approaches for dimensionality reduction:

- **Feature extraction: Transforming the existing features into a lower dimensional space.**
- **Feature selection: Selecting a subset of the existing features.**

Feature Section searches for a subset that minimizes some cost function (e.g. test error)

Feature Subset Selection

Given a set of n features, the goal of feature selection is to select a subset of d features (d < n) in order to minimize the error.

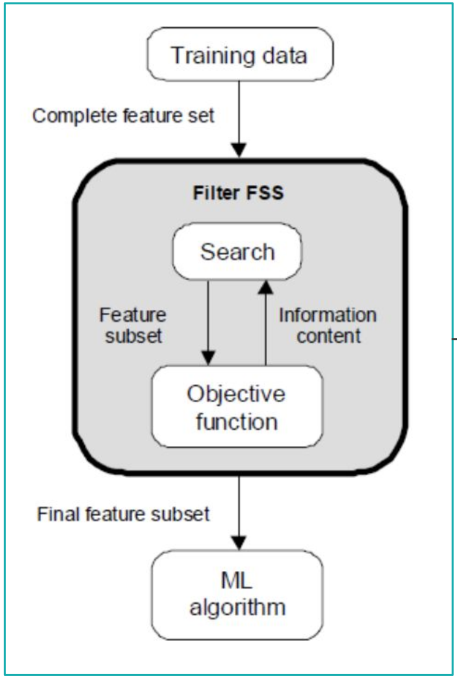


Feature Subset Selection requires:

- A search strategy to select candidate subsets.
- An objective function to evaluate these candidates.

Basic Process

The objective function evaluates subsets and returns a measure of their "goodness".



<u>Step 1</u>: Search the space of possible feature subsets.

<u>Step 2:</u> Pick the sub set that is optimal or near-optimal with respect to some objective function

Search Strategy

Assuming n features, an exhaustive search would require:

1. Examining all possible subsets of size d. → $\binom{n}{d}$

2. Selecting the subset that performs the best according to the Objective function.

The number of subsets grows combinatorially, making exhaustive search impractical.

Naïve Search

Step 1: Sort the given n features in order of their probability of correct recognition.

Step 2: Select the top d features from this sorted list.

Disadvantage:

- Correlation among features is not considered.
- The best pair of features may not even contain the best individual feature.

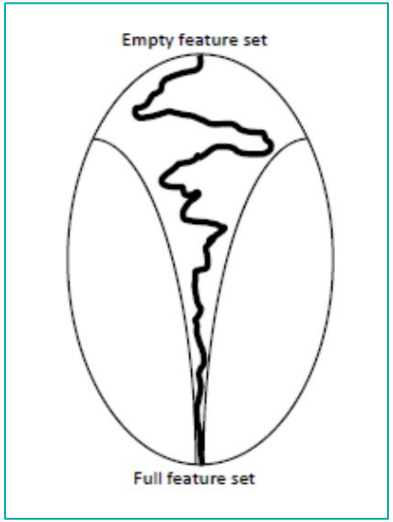
Sequential forward selection (SFS)

Step 1: First, the best single feature is selected (i.e., using some objective function).

Step 2: Then, pairs of features are formed using one of the remaining features and this best feature, and the best pair is selected.

Step 3: Next, triplets of features are formed using one of the remaining features and these two best features, and the best triplet is selected.

This procedure continues until a predefined number of features are selected.



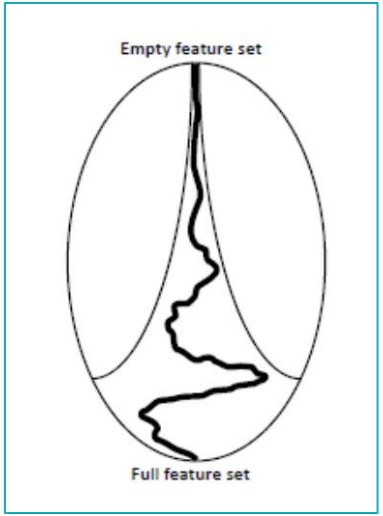
Sequential backward selection (SBS)

Step 1: First, the objective function is computed for all n features.

Step 2: Then, each feature is deleted one at a time, the objective function is computed for all subsets with n-1 features, and the worst feature is discarded.

Step 3: Next, each feature among the remaining n-1 is deleted one at a time, and the worst feature is discarded to form a subset with n-2 features.

This procedure continues until a predefined number of features are selected.



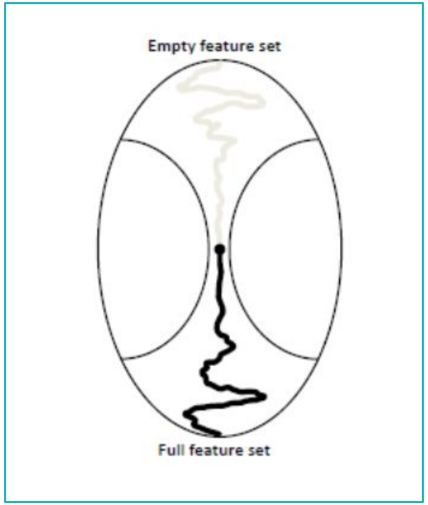
Bidirectional Search (BDS)

BDS applies SFS and SBS simultaneously:

SFS is performed from the empty set & SBS is performed from the full set.

To guarantee that SFS and SBS converge to the same solution:

- Features already selected by SFS are not removed by SBS.
- Features already removed by SBS are not added by SFS.



Limitations of SFS and SBS

The main limitation of SFS is that it is unable to remove features that become non useful after the addition of other features.

The main limitation of SBS is its inability to reevaluate the usefulness of a feature after it has been discarded.

Objective Function Types

Filter Objective Functions:

- Evaluation is independent of the classification algorithm.
- The objective function evaluates feature subsets by their information content, typically interclass distance, statistical dependence or information-theoretic measures.

Wrapper Objective Functions:

- Evaluation uses criteria related to the classification algorithm.
- The objective function depends on predictive accuracy (recognition rate on test data) by statistical resampling or cross- validation.

Filter Methods - Advantages and Limitations

Advantages:

- Fast execution: Filters generally involve a non-iterative computation on the dataset, which can execute much faster than a classifier training session.

- Generality: Since filters evaluate the intrinsic properties of the data, rather than their interactions with a particular classifier, their results exhibit more generality: the solution will be "good" for a larger family of classifiers

Disadvantages:

- Tendency to select large subsets: The filter tends to select the full feature set as the optimal solution. This forces the user to select an arbitrary cutoff on the number of features to be selected.

Wrapper Methods - Advantages and Limitations

Advantages:

- Accuracy: Wrappers generally achieve better recognition rates than filters since they are tuned to the specific interactions between the classifier and the dataset.

- Ability to generalize: Wrappers have a mechanism to avoid overfitting, since they typically use cross-validation measures of predictive accuracy.

Disadvantages:

- Slow execution: since the wrapper must train a classifier for each feature subset the method can become infeasible for computationally intensive methods.

- Lack of generality: The "optimal" feature subset will be specific to the classifier under consideration and lacks generality

Much obliged.

TECH I.S.