

FEATURE SELECTION METHODS

FILTER METHODS | OVERVIEW



FILTER METHODS



- Select variables independently of the machine learning algorithm
 - Rely **only** on the characteristics of the data (of the variables)
-
- Model agnostic
 - Fast computation

FILTER METHODS | UNIVARIATE



Two step procedure:

- Rank features according to a certain criteria
 - Each feature is ranked independently of the feature space
 - Select the highest ranking features
- May select redundant variables because they **do not consider the relationships** between features

FILTER METHODS | RANKING CRITERIA



Feature scores on various statistical tests:

- Chi-square | Fisher Score
- Univariate parametric tests (anova)
- Mutual information
- Variance
 - Constant features
 - Quasi-constant features

FILTER METHODS | MULTIVARIATE



- Handle redundant feature
 - Duplicated features
 - Correlated features
-
- Simple yet powerful methods to quickly remove irrelevant and redundant features
 - First step in feature selection procedures

FILTER METHODS



- Quick dataset screening for irrelevant features
- Quick removal of redundant features
 - Constant features
 - Duplicated features
 - Correlated features