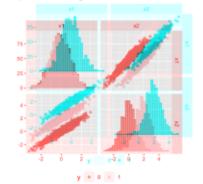
Introduction to Machine Learning

Feature Selection: Filter Methods

Feature Selection: Filter Methods

(Examples and Caveats)

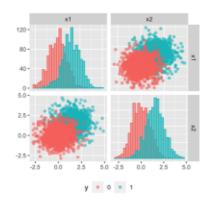


earning goals Learning goals

- Understand how filter method
 - Understand how filter methods can be misleading
 - Understand how filters can be
 - Understand how filters can be applied and tuned



FILTER METHODS CAN BE MISLEADING



 ρ_{ACC} of log. reg. classifier with:

feature x₁: 0.76

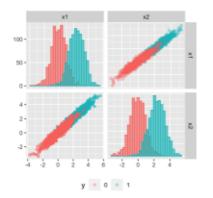
feature x₂: 0.78

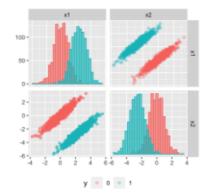
both features: 0.85



Information gain from presumably redundant variables. 2 class problem with indep features. Each class has Gaussian distribution with no covariance. While filter methods suggest redundancy, combination of both vars yields improvement, showing indep vars are not truly redundant. For further details, see Calyon and Elisage 2003

FILTER METHODS CAN BE MISLEADING /2

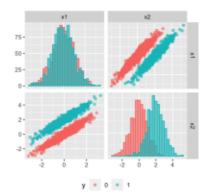


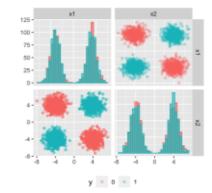




Intra-class covariance. In projection onto axes, distribution two variables are same as before a Lefth Class conditional distribuve high cov. in direction of the line of two class centers. Righth Class conditional distribuve high cov. in direction perpendicular to line of two class centers. Better separation by using both varsation gain is obtained by using both variables.

FILTER METHODS CAN BE MISLEADING /3







Variable useless by itself can be useful together with others. Left: One var has completely overlapping class conditional densities. Still, jointly with other variable separability can be improved. Right: XOR-like chessboard problem. Classes consist of "clumps" s.t. projection on the axes yields overlapping densities. Single vars have no separation power, only used together.

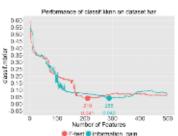
USING FILTER METHODS

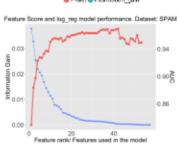
- Calculate filter score for each feature x_j
- Rank features according to score values.
- Choose p best features
- Train model on p

 best features.

How to choose \tilde{p} ?

- Could be prescribed by application
- Eyeball estimation: read from filter plots
- Treat as hyperparameter and tune in a pipeline, based on resampling







USING FILTER METHODS /2

Advantages:

- Easy to calculate
- Typically scales well with the number of features p
- Generally interpretable
- Model-agnostic

Disadvantages:

- Univariate analyses may ignore multivariate dependencies
- Redundant features will have similar weights
- Ignores the learning algorithm

