

Method description

Thomas Navin Lal and Olivier Chapelle

{navin.lal, olivier.chapelle}@tuebingen.mpg.de

Max Planck Institute for Biological Cybernetics,
Tübingen, Germany



approach



- Preprocessing: data was rescaled
- Features are ranked using the Fisher Criterion:
$$R(f) = \frac{(\mu(f^+) - \mu(f^-))^2}{\text{var}(f^+) + \text{var}(f^-)}$$
- The number of features was determined by cross validation
- Classification was done by SVMs (linear or RBF)
- Heuristic used to find σ : Let t_i denote the distance of point x_i to the set formed by the points of the other class. σ is set to the mean of t_i .



datasets



- Arcene: rescaled. 4700 features. SVM linear $C=\text{inf}$
- Dexter: 3700 features. SVM linear $C=\text{inf}$
- Dorothea: rescaled. 1000 features. SVM $\sigma=0.11$ $C=10$
(+balanced)
- Gisette: rescaled. 700 features SVM $\sigma=0.38$ $C=\text{inf}$
- Madelon: rescaled 20 features. SVM $\sigma=0.01$ $C=\text{inf}$