

Nameless

Feature Selection Challenge Attempt

By

Ran Gilad-Bachrach and Amir Navot

Overview

- In most cases we have used standard “out of the box” algorithms
- Obvious modifications for balanced error were done
- A novel feature selection algorithm was introduced ([distBased](#))
- Over fit was probably done by running over too many algorithms with too many parameters

Classification Method

- SVM
 - We have used the [SVM toolbox](#) by **Gavin Cawley** (University of East Anglia, England)
- Naïve Bayes
 - Good-Turing zero correction
- Perceptron
 - Aggressive version ([Crammer et al.](#))

Feature Selection Methods

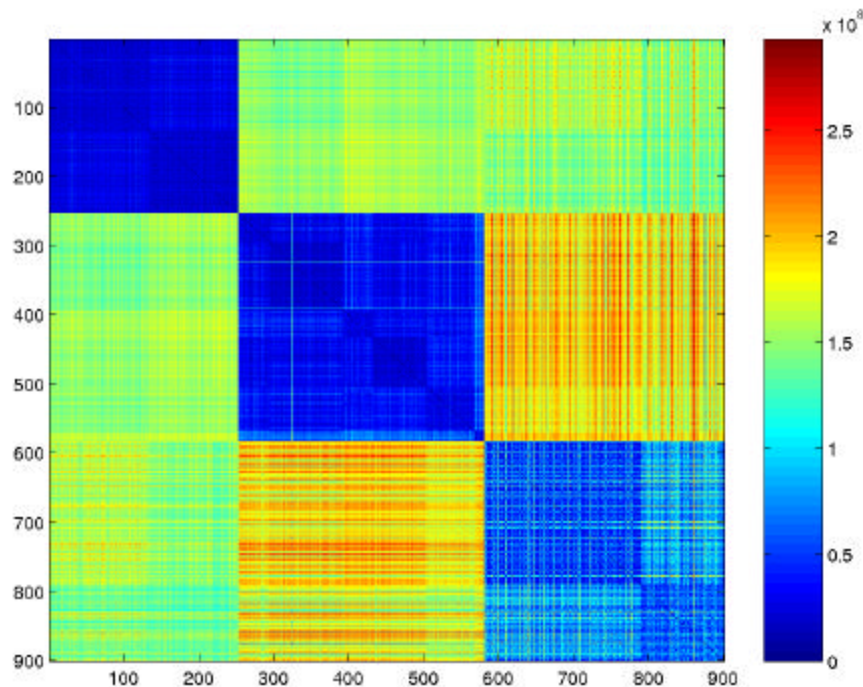
- MI1
 - features are scored by the mutual information between the feature value and the labels
 - Non binary data, was compared to the median
- MI2
 - same as MI1 while zero valued featured are assumed to be sleeping

Feature Selection Methods – Cont.

- DistBased
 - [CGNT02](#) defined the proper margin for prototype based algorithms (Nearest Neighbor, LVQ, SVM-RBF)
 - The margin of an instance is the difference between the distance to the closest negative prototype and the closest positive prototype
 - We selected features that maximizes this margin

Arcene - Observation

- The data has a clear hierarchical structure, which can be revealed by clustering
- The figure shows the mutual distance between instances



- The instances were reordered by k-means

Arcene – Algorithm

- Normalization: The maximum absolute value of each feature was set to 1
- Representation: PCA
- Feature selection: distBased. 81 principal components were used.
- Classification: SVM
 - Kernel: rbf(0.005)
 - C=8

Gisette - Algorithm

- Normalization: The maximum absolute value of each feature was set to 1
- Feature selection: MI1
- Classification: aggressive perceptron with a limit set to 600 (i.e. we require that $y(w \cdot x) > 600$ for each (x,y) in the training set).

Dexter - Algorithm

- Normalization: none
- Feature selection: MI1
- Classification: Transductive SVM
 - Kernel: linear
 - $C=10$
 - 3 transduction rounds with addition of 15% of the unlabeled sample in each round.

Dorothea - Algorithm

- Normalization: none
- Feature selection: MI2
- Classification:
 - Naïve Bayes
 - Good Turing Zero Correction

Madelon - Algorithm

- Normalization: The maximum absolute value of each feature was set to 1
- Feature selection: distBased
- Classification: Transductive SVM
 - Kernel: rbf(50)
 - $C=5$
 - 13 transduction rounds. in each round 10% of the unlabeled data was added.