# Feature Extraction Results

Markus Uhr

## Background

The students had the opportunity to experiment with the classification problems of the NIPS 2003 Feature Selection Challenge. Techniques and methods presented during the lecture could be applied and practised. The goal was to improve on a given baseline method and try to reach the performance of the contest winners.

A Matlab toolbox for machine learning was provided for easy testing and to lower the work to implement own ideas.

## Datasets

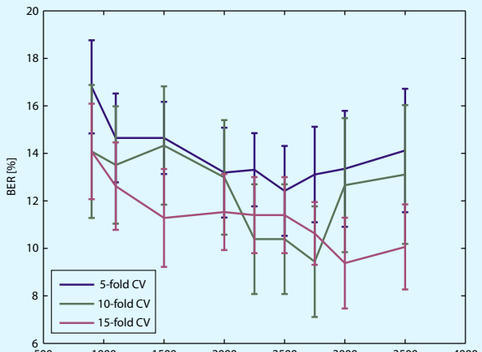
| Dataset | Size (MB) | Type | Features | Training | Validation | Test |
|---|---|---|---|---|---|---|
| Arcene | 8.7 | dense | 10 000 | 100 | 100 | 700 |
| Gisette | 22.5 | dense | 5000 | 6000 | 1000 | 6500 |
| Dexter | 0.9 | sparse | 20 000 | 300 | 300 | 2000 |
| Dorothea | 4.7 | sparse | 100 000 | 800 | 350 | 800 |
| Madelon | 2.9 | dense | 500 | 2000 | 600 | 1800 |

## Methods

- Learning objects from CLOP www.modelselect.inf.ethz.ch/models.php
- Performance measured with BER (balanced error rate), i.e. the averaged positive and negative class error rates
- Baseline model: consisting of given CLOP objects with BER0 performance

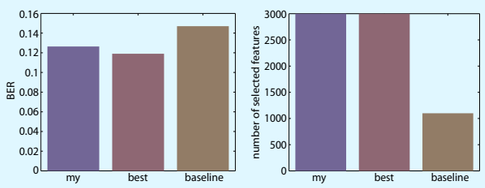
## Arcene

***cancer diagnosis***




**Model**

1. Feature standardization
2. s2n filter (f_max=3000)
4. Feature normalization
5. SVM (coef=1, deg=2, gam=1, shrink=0.1)




**Variants**

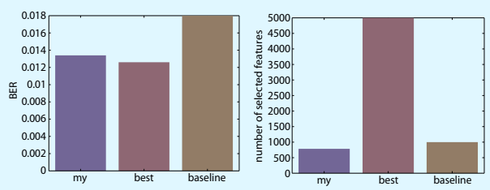
- Cluster data and train each cluster individually
- Boosting (stumps) classifier

## Gisette

***digit recognition***

**Model**

1. Skeletonization filter
2. Low-pass filter (Mitchell 9x9)
3. Scale and translation normalization
4. Feature normalization
5. SVM (coef=1, deg=3, gam=0, shrink=1)

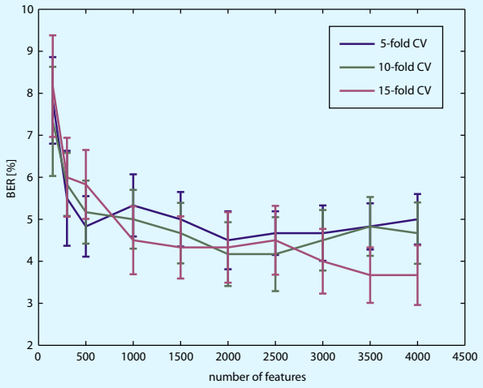



**Variants**

- Different low-pass filters (e.g. Gaussian, Exponential, Lancoz)
- Boosting classifier with stumps (about the same performance as SVM with deg=1)
- Rotation invariant features based on Zernike polynomials

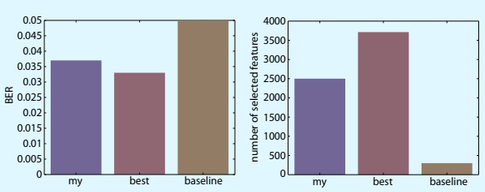
## Dexter

***text categorization***




**Model**

1. s2n filter (f_max=2500)
2. Feature normalization
3. SVM (coef=0, deg=1, gam=0, shrink=0.5)

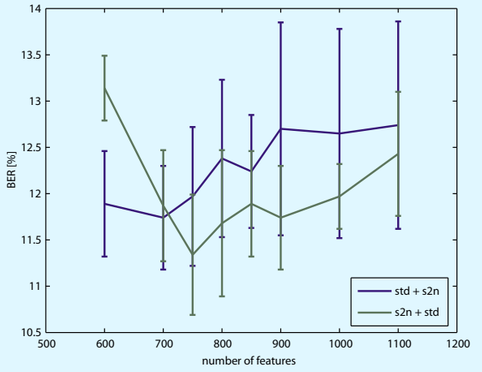



**Variants**

- none

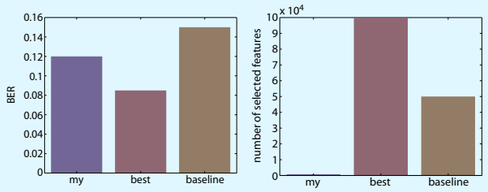
## Dorothea

***drug discovery***




**Model**

1. TP filter (true positives ranking creteria, 2000 features selected)
2. Feature standardization
3. s2n filter (f_max=750)
4. Naive bayes classifier
5. Bias reduction (post-processing)

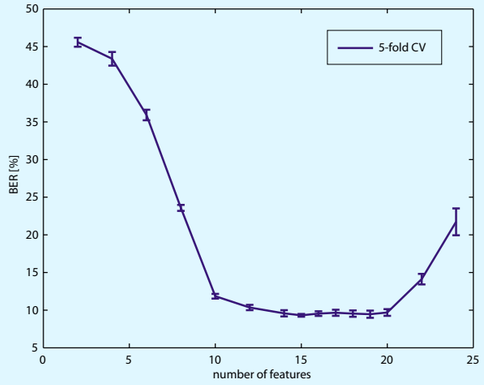



**Variants**

- Different filters in the spirit of TP (using also false negatives)
- Swap feature standardization and s2n filter

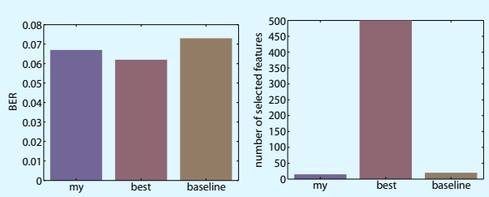
## Madelon

***artificial data***




**Model**

1. Feature selection with probe method (2000 probes, Relief method for feature ranking, 15 features selected)
2. Feature standardization
3. SVM (coef=1, deg=0, gam=1, shrink=1)




**Variants**

- Boosting (stumps) classifier

## Conclusion

In the case of Gisette where the origin and the meaning of the features is known and can be interpreted, well directed and specific preprocessing could increase the classification rate considerably. Also, visualization of the results gave a better understanding of the problem.

In the other cases, use of standard preprocessing tools (e.g. feature standardization, simple filters) and support vector classifiers lead to good classification rates when exploring their parameter space. Despite the success in these cases, a deeper understanding of the problem is not given.

## References


- NIPS Challenge Homepage http://www.nipsfsc.ecs.soton.ac.uk/
- CLOP Toolbox www.modelselect.inf.ethz.ch/models.php
- Course Book *Feature Extraction, Fundamentals and Applications*, I. Guyon et al, to appear in Springer.