Feature Extraction Course – Poster Presentation

Patrick Pletscher

Abstract

We present the results of our submission for the Nips 2003 Feature Extraction challenge, which was organized by I. Guyon et al.

1 Background

Two possible drawbacks when learning on entire dataset:

- Computational load too high when using a complicated learner on the whole training data set.
- Statistical stability of classifier is affected by "learning" garbage-features.

It is thus often advisable to divide the learning in two steps:

1. Find relevant features.
2. Do the learning only on the relevant features.

2 Contest

- Five datasets from various areas.
- Two-class classification problems.

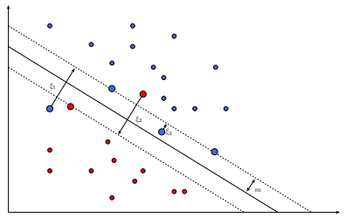
Challenge participants were classified according to their balanced error rate (BER) on the test set.

3 Some Observations

- Support Vector Machines are usually good classifiers to get started with.
- Often the joice of filter is not too crucial.

4 Support Vector Machines

As we use support vector machines all over again, we think it is useful to review some fundamentals of the SVMs, such that we have them handy.



Soft margin SVM with 2-norm regularization,

$$\text{minimize} \quad \|\boldsymbol{w}\|^2 + C \sum_j \xi_j^2$$

$$\text{subject to} \quad y_j(\boldsymbol{w} \cdot \boldsymbol{x}_j + w_0) \geq (1 - \xi_j)$$

Radial basis kernel for the "kernel-trick":

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|}{2\sigma^2}\right)$$

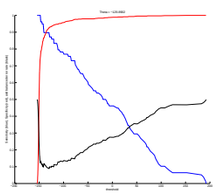
5 Results

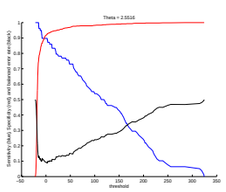
5.1 Dorothea: Thrombin dataset

Learning on training and validation data set.

```
my_model=chain({TP('f_max=2000'),
probe(relief,{'f_max=700'}), standardize, naive,
bias});
```

Importance of bias adjustment (left: unbiased, right: biased).



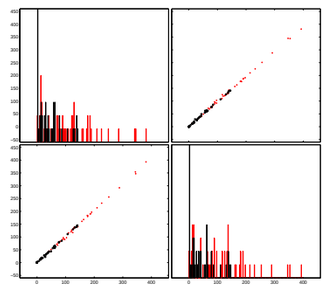


5.2 Arcene: cancer diagnosis

Learning on training and validation data set.

```
my_svc=svc({'coef0=1', 'degree=3', 'gamma=0',
'shrinkage=0.1'});
my_model=chain({standardize, Ftest('f_max=2740'),
normalize, my_svc});
```

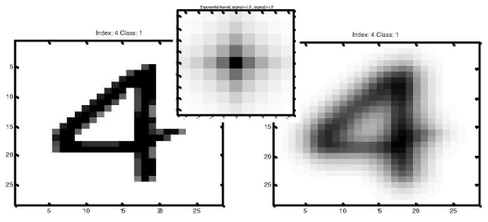
Not all filter methods perform equally good!



5.3 Gisette: handwritten digits

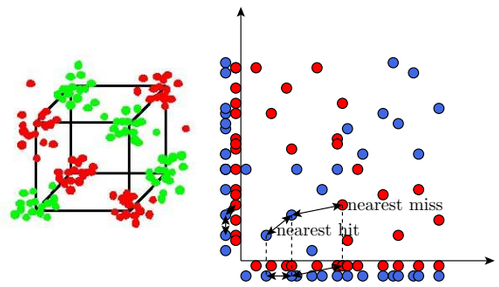
Construct new features by convolution with an exponential kernel.

```
my_classif=svc({'coef0=1', 'degree=4', 'gamma=0',
'shrinkage=0.03'});
my_model=chain({convolve(exp_ker({'dim1=9',
'dim2=9', 'sigma1=1.7', 'sigma2=1.7'})),
s2n('f_max=4500'), normalize, my_classif});
```



5.4 Madelon: random data

Learning on training and validation data set.



Data is inherently non-linear. Don't punish non-linear functions too much (shrinkage of SVM small) and use a non-linear filter method (relief).

```
my_classif=svc({'coef0=1', 'degree=0', 'gamma=0.3',
'shrinkage=0.3'});
my_model=chain({probe(relief,{'p_num=2000',
'f_max=20'}), standardize, my_classif});
```

5.5 Dexter: filters text

NEW YORK, October 2, 2001 – Instinent Group Incorporated (Nasday: INET), the world's largest electronic agency securities broker, today announced that it has completed the acquisition of ProTrader Group LP, a provider of advanced trading technologies and electronic brokerage services primarily for retail active traders and hedge funds. The acquisition excludes ProTrader's properitary trading business. ProTrader's 2000 annual revenues exceeded $83 million.

```
my_classif=svc({'coef0=1', 'degree=1', 'gamma=0',
'shrinkage=0.5'});
my_model=chain({s2n('f_max=2000'), normalize,
my_classif});
```

6 Overall results

Comparison to the best challenge entry for each data set.