Selected Topics in Applied Machine Learning: An integrating view on data analysis and learning algorithms

ESSLLI '2015
Barcelona, Spain

http://ufal.mff.cuni.cz/esslli2015

Barbora Hladká
hladka@ufal.mff.cuni.cz

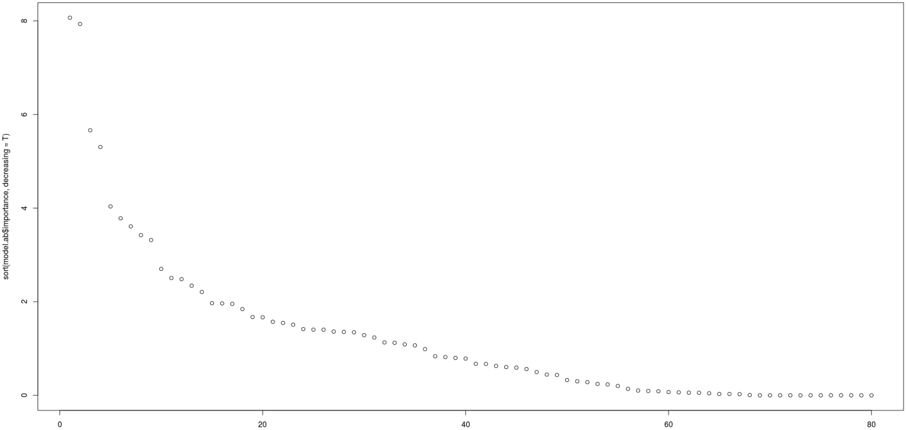
Martin Holub
holub@ufal.mff.cuni.cz

Charles University in Prague,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics

Block 5.1
Practical methods for feature selection

- **Filters and wrappers**
- **Variable importance produced by ensembles**
- **Feature selection by Lasso**
- **SVM-RFE – Recursive Feature Elimination**

Variable importance (AdaBoost) – *cry*



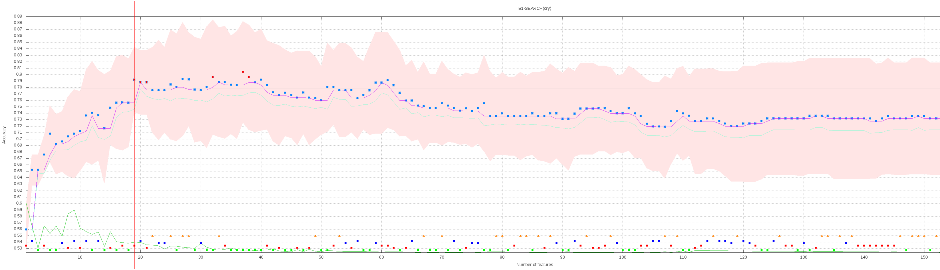
SVM-RFE feature selection algorithm

Algorithm 2 Recursive feature elimination using the SVM learner with cross-validated optimization of the SVM parameter *cost* in each iteration step.

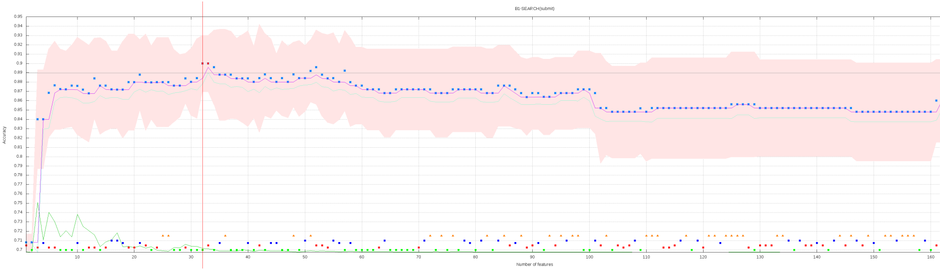
Input: Training data set and the initial feature set
Output: The best SVM classifier $\mathrm{M}_{\max}$ and the corresponding feature subset $\mathrm{S}_{\max}$

$K \leftarrow$ *the initial feature set size*
$S_K \leftarrow$ *the initial feature set*
for $k \leftarrow K$ **downto** 1 **do**
  learn a linear SVM model using the feature set S_k *and tune its parameter* cost
  $\mathrm{M}_k \leftarrow$ *the best tuned linear SVM model using the feature set* S_k
  $f_{\text{worst}} \leftarrow$ *the least useful feature in the model* M_k
  $S_{k-1} \leftarrow S_k \setminus \{f_{\text{worst}}\}$
end for
$\mathrm{M}_{\max} \leftarrow$ *choose the best model from* $\{\mathrm{M}_i\}_{i=1}^{K}$
$\mathrm{S}_{\max} \leftarrow$ *the best feature subset corresponding to the best model* $\mathrm{M}_{\max}$

SVM-RFE – *cry*



SVM-RFE – *submit*



Block 5.2
Summarizing remarks on VPR task
– model assesment and selection

Goal

A complex comparison of competing models trained for the VPR task

- **Model** = method + set of features + learning parameters
  – feature set may be considered as a parameter of the model (!)

- Model flexibility is model's ability to fit the data well

- Higher flexibility implies higher complexity **but not vice versa**

Outline

- **Concluding remarks on methods for reducing the variance**
  – ensembles vs. feature selection vs. regularization
- **The "Bayes classifier" – the limit of the test error**
- **A bootstrap method for estimating the generalization error**
- **More complex comparison of the developed VPR classifiers**
  – confidence intervals, an indicator of the model variance
- **Concluding remarks & concluding questions**

Model assessment and model selection

Assessment – a dictionary definition

- *Assessment means the evaluation or estimation of the nature, quality, or ability of someone or something.*

Model assessment and model selection

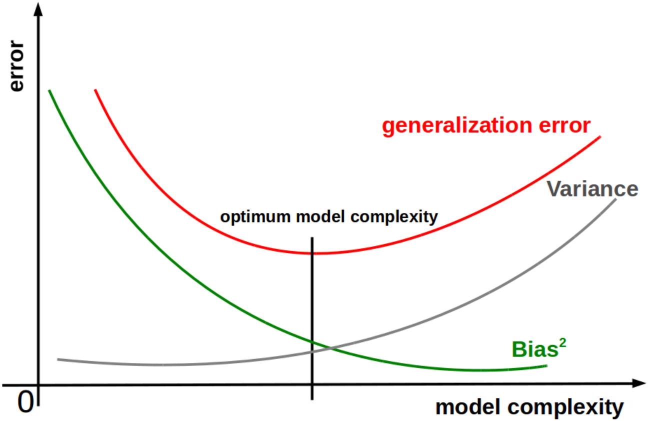
Assessment – a dictionary definition

- *Assessment means the evaluation or estimation of the nature, quality, or ability of someone or something.*

- **Model assessment**
  – the process of evaluating a model's performance
- **Model selection**
  – the process of selecting the proper level of flexibility

Model assessment and model selection

Finding a model that minimizes generalization error



Model assessment and model selection

In a linguistic research, we are also interested in the interpretability of the model, namely in

- recognizing/discovering important features
- error analysis

Unstable classifiers and ensembles

Leo Breiman:
"Instability (of base learners) is an essential ingredient for bagging or arcing to improve accuracy."

Unstable classifiers and ensembles

Leo Breiman:
"Instability (of base learners) is an essential ingredient for bagging or arcing to improve accuracy."

- Unstable classifiers are characterized by high variance
- Decision trees are especially suitable for building ensembles because
  - they are extremely flexible to fit "any data",
    i.e they can have very low bias
  - they are unstable, i.e. they have high variance

Ensembles and feature selection and regularization

How and why do the techniques for decreasing variance work?

Ensembles and feature selection and regularization

How and why do the techniques for decreasing variance work?

- feature selection and regularization decrease model complexity

Ensembles and feature selection and regularization

How and why do the techniques for decreasing variance work?

- feature selection and regularization decrease model complexity
- bagging-based ensembles average a large set of low correlated results

Ensembles and feature selection and regularization

How and why do the techniques for decreasing variance work?

- feature selection and regularization decrease model complexity
- bagging-based ensembles average a large set of low correlated results
- AdaBoost decreases bias in the early iterations
  – it decreases variance as well, namely in later iterations

What is the lowest possible error rate

Bayes classifier assigns each example to the most likely class, given its feature values

$$\hat{y} = max_y \Pr(y \,|\, \mathbf{x})$$

The Bayes classifier produces the lowest possible test error rate, so called **Bayes error rate**

$$1 - \mathrm{E}\left(max_y \Pr(y \,|\, \mathbf{x})\right)$$

What is the lowest possible error rate

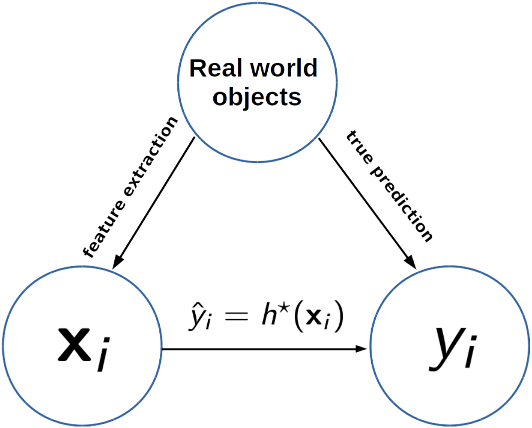
Identical feature vectors?

What is the lowest possible error rate

Identical feature vectors?

- Get the same feature vectors
- How many of them have the same target value?

What is the lowest possible error rate



Competing models for VPR

- **Confidence intervals**

Generalization error estimation by bootstrapping

- Suppose a development data of n examples
- Train a model on the data
- Test the model on the data
- Get training error = optimistic error e_l
- Repeat 200 times
  - Randomly select n examples with replacement and train a model
    on average, 63.2% of the original sample
  - Test the model on the examples not used in the training
    on average, 36.8% of the orginal sample
  - Get test error
- Get mean test error = pesimistic error e_o
- **generalization error estimation =** $0.368 * e_l + 0.632 * e_o$

Final look at our competing models

Overview of all models developed

Farewell

It's Friday afternoon! No more slides, no more lectures!

Farewell

It's Friday afternoon! No more slides, no more lectures!

- **It was a pleasure for us to be here with you.**
- **We are glad that we could teach you something.**
- **You were the bright audience.**

Thank you!