Universität Bielefeld

Cognitive Interaction Technology Center of Excellence

How to visualize a classifier?

A. Schulz, A. Gisbrecht, K. Bunte and B. Hammer

{aschulz|agisbrec|kbunte|bhammer}@techfak.uni-bielefeld.de

University of Bielefeld

28.08.2012

Quality assessment

How is the result of a classifier usually accessed?

Quality assessment

How is the result of a classifier usually accessed?

89.76%

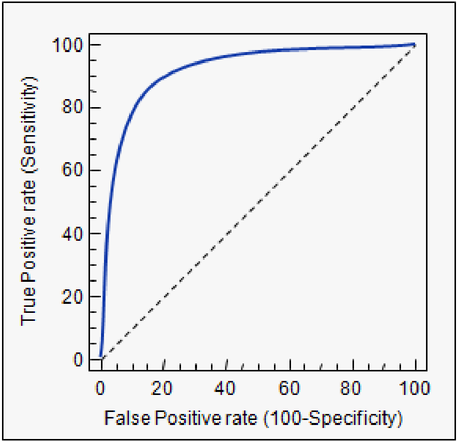
Quality assessment

How is the result of a classifier usually accessed?

$$89.76\%\,(0.05)$$

Quality assessment

How is the result of a classifier usually accessed?



Quality assessment

How is the result of a classifier usually accessed?

Table: Confusion matrix

	1	2	3	4	5	$\sum$
1	176	10	12	7	2	207
2	1	57	11	9	3	81
3	18	25	43	31	10	127
4	1	5	23	127	4	160
5	4	3	11	26	131	175
$\sum$	200	100	100	200	150	750
class-wise accuracy of estimation in %						
	88.00	57.00	43.00	63.50	87.33	

What else could be interesting?

More information about the result:

- Which samples near to the decision boundary?
- Which samples are classified incorrectly and “how much”?
- How smooth is the decision boundary?
- How are the modes of the classes structured?
- How are errors distributed?

Where is the Difficulty?

Class borders are

- Often non linear
- Often not given in an explicit functional form (e.g. SVM)
- High-dimensional $\rightarrow$ non feasible for visualization

Our approach

1. Project the trained data into 2D
   DR method (should map relative positions of classes faithfully)

Our approach

1. Project the trained data into 2D
   DR method (should map relative positions of classes faithfully)
2. Sample the 2D data space
   much less effort than sampling in the high-dimensional space

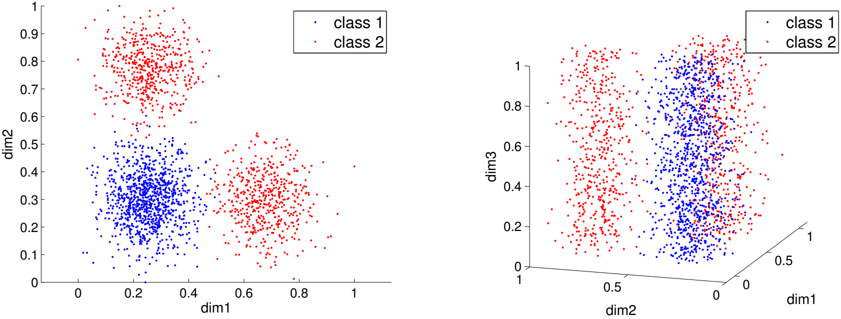
Our approach

1. Project the trained data into 2D
   DR method (should map relative positions of classes faithfully)
2. Sample the 2D data space
   much less effort than sampling in the high-dimensional space
3. Project the samples up to the original space

Our approach

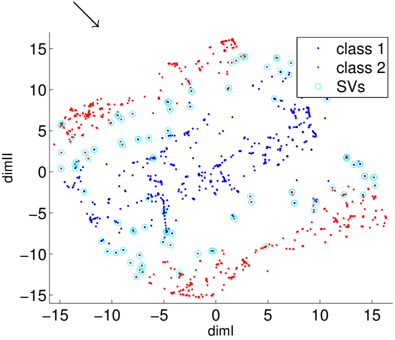
1. Project the trained data into 2D
   DR method (should map relative positions of classes faithfully)
2. Sample the 2D data space
   much less effort than sampling in the high-dimensional space
3. Project the samples up to the original space
4. Classify them

An illustration: ”high”-dimensional data



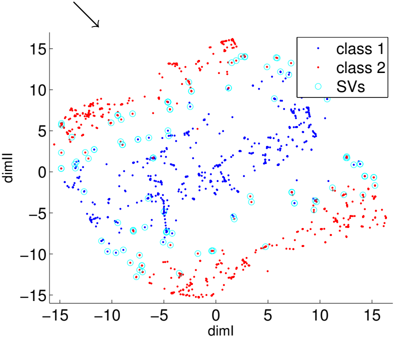
An illustration: dimensionality reduction

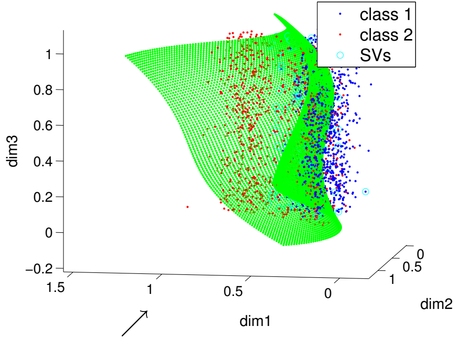
Train the classifier and project data to 2D



An illustration: dimensionality reduction

Train the classifier and project data to 2D

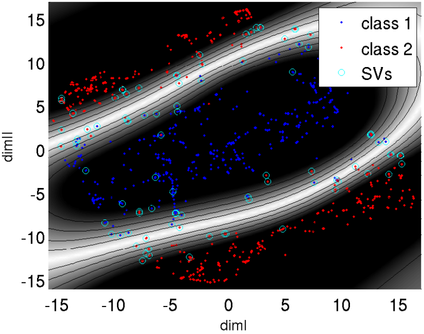




Sample in 2D and project up

An illustration: border visualization

- contours code the distance from the class boundaries



Inverting the dimensionality reduction

Assume n training data $\boldsymbol{x}_i \in \mathbb{R}^D$ accompanied by labels $l_i \in L$

- Use a nonlinear dimension reduction method to produce a low-dimensional embedding $p(\boldsymbol{x}_i) = \boldsymbol{y}_i \in \mathbb{R}^2$

Inverting the dimensionality reduction

Assume n training data $\boldsymbol{x}_i \in \mathbb{R}^D$ accompanied by labels $l_i \in L$

- Use a nonlinear dimension reduction method to produce a low-dimensional embedding $p(\boldsymbol{x}_i) = \boldsymbol{y}_i \in \mathbb{R}^2$
- Define a kernel mapping

$$\boldsymbol{x}_j \approx p^{-1}(\boldsymbol{y}_j) = \frac{\sum_i \alpha_i k_i(\boldsymbol{y}_i, \boldsymbol{y}_j)}{\sum_i k_i(\boldsymbol{y}_i, \boldsymbol{y}_j)} = \mathbf{AK}$$

with parameters α, the kernel and possible kernel parameters

Inverting the dimensionality reduction

Assume n training data $\boldsymbol{x}_i \in \mathbb{R}^D$ accompanied by labels $l_i \in L$

- Use a nonlinear dimension reduction method to produce a low-dimensional embedding $p(\boldsymbol{x}_i) = \boldsymbol{y}_i \in \mathbb{R}^2$
- Define a kernel mapping

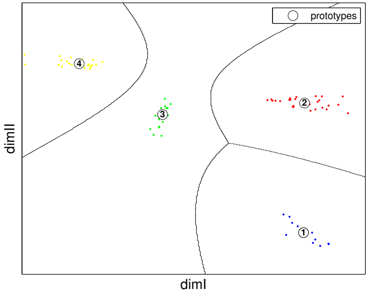
$$\boldsymbol{x}_j \approx p^{-1}(\boldsymbol{y}_j) = \frac{\sum_i \alpha_i k_i(\boldsymbol{y}_i, \boldsymbol{y}_j)}{\sum_i k_i(\boldsymbol{y}_i, \boldsymbol{y}_j)} = \mathbf{AK}$$

  with parameters α, the kernel and possible kernel parameters
- Find the least squares solution for α of the reconstruction error

$$E = \sum_j \left\| \boldsymbol{x}_j - p^{-1}(\boldsymbol{y}_j) \right\|^2$$

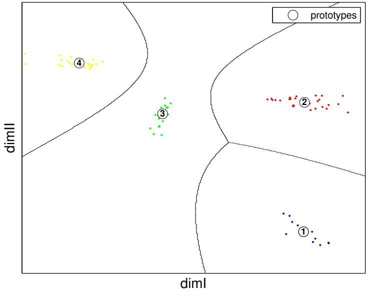
Application

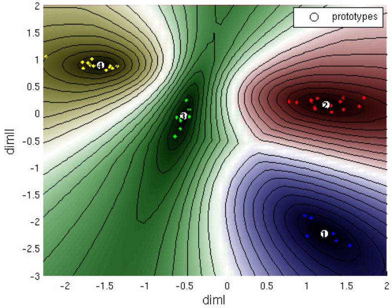
- Gene Expression Data Set (Ground truth available)
- Visualization via LLiRaM LVQ [Bunte, Dissertation]



Application

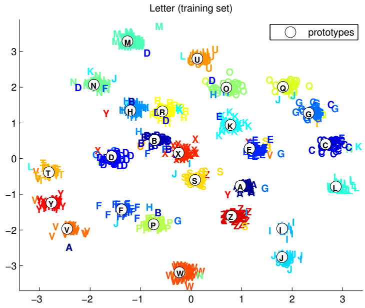
- Gene Expression Data Set (Ground truth available)
- Visualization via LLiRaM LVQ [Bunte, Dissertation]





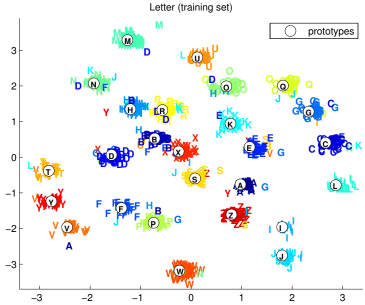
Application

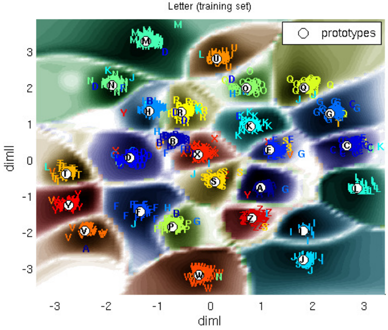
- Letter Recognition Data Set
- Visualization via LiRaM LVQ + t-sne [Bunte et.al 2012]



Application

- Letter Recognition Data Set
- Visualization via LiRaM LVQ + t-sne [Bunte et.al 2012]





Conclusions & Future work

Conclusion:

- The visualization of classification boundaries is possible
- First experiments showed promising results

Future Work:

- What if certain dimensions are more relevant for classification? (Fisher information, classifier)
- Learning of hyper-parameters

Thank You For Your Attention!