Training of Radial Basis Function Classifiers With Resilient Propagation and Variational Bayesian Inference

D. Fisch, B. Sick

{fisch, sick}@fim.uni-passau.de

UNIVERSITY OF PASSAU

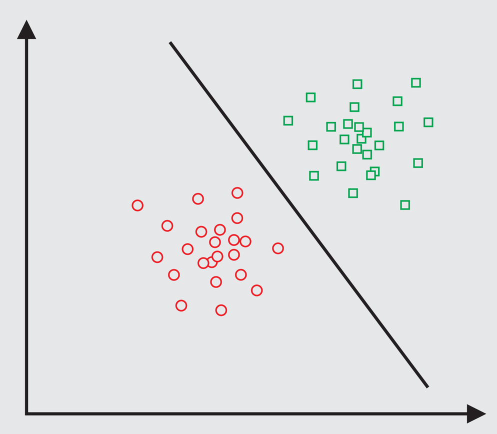
Faculty of Informatics and Mathematics

Institute for Computer Architecture

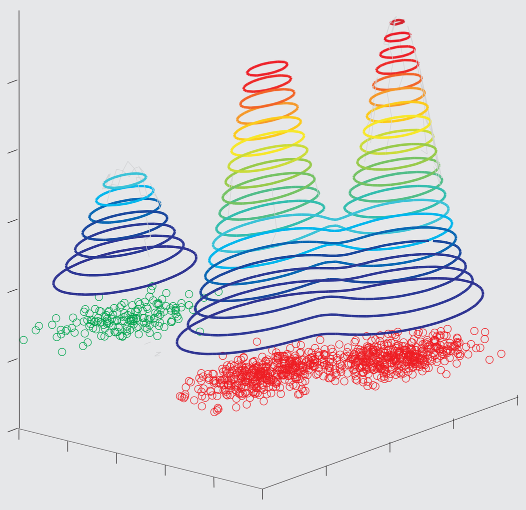
Computationally Intelligent Systems

Types of Classifiers

- **Discriminative Classifiers:**
  Use discriminant function $f(x) \rightarrow c$ to directly map an input sample x to a class c.



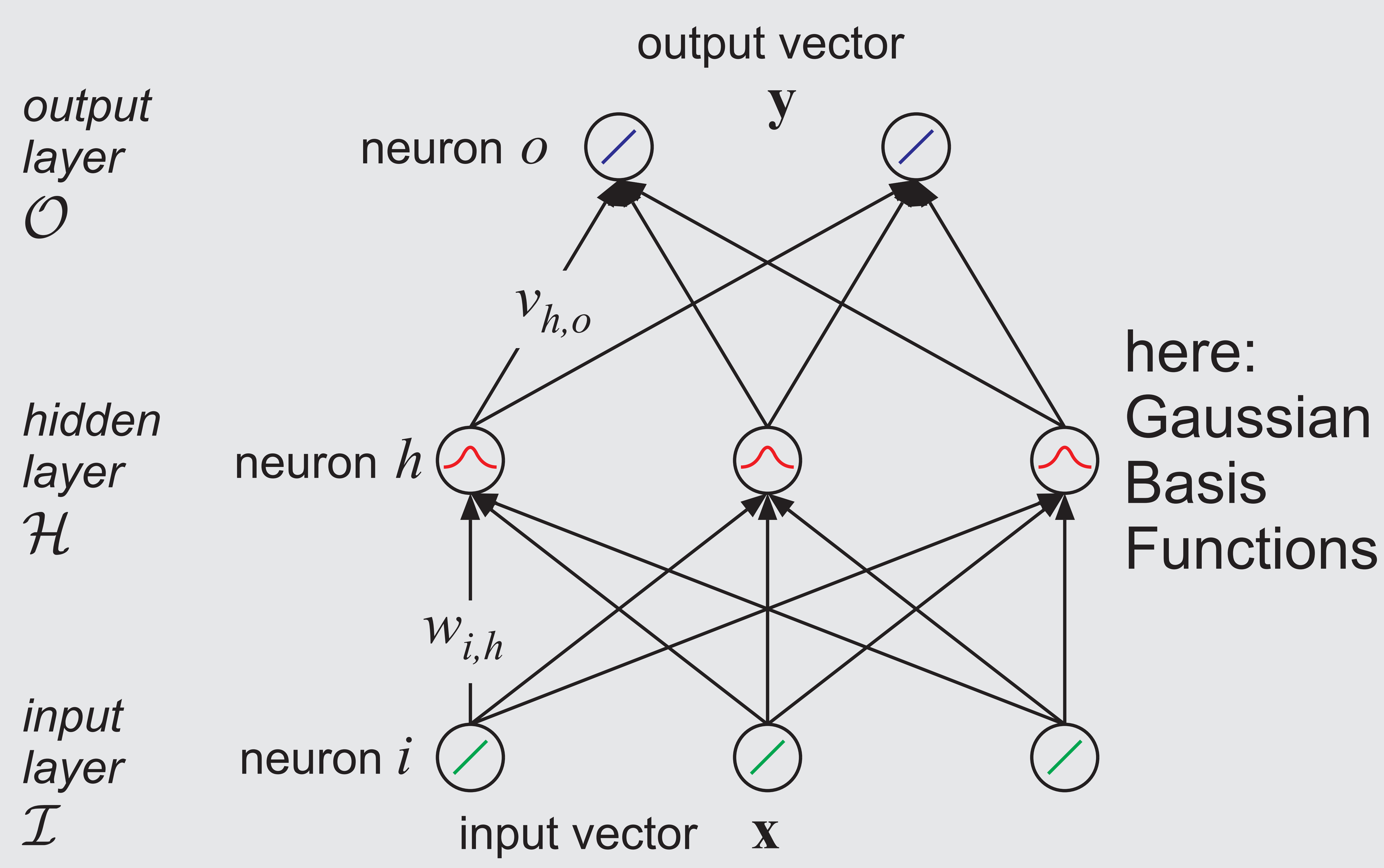
- **Generative Classifiers:**
  Try to model the data generating processes, e.g., by means of a probabilistic distribution $p(x)$.



Advantages of Generative Classifiers

- Support of **loss functions** to minimize the risk of wrong actions following from a classification decision.
- Possibility to define a **rejection criterion** to refuse a decision if uncertainty is too high.
- If certain types of density functions are used: Possibility to **extract symbolic rules** that can be understood by human domain experts.
- Support of **situation-awareness**: Possibility to detect changes in the environment (new processes emerge, existing processes become obsolete) and to react accordingly.

Radial Basis Function Networks (RBF)

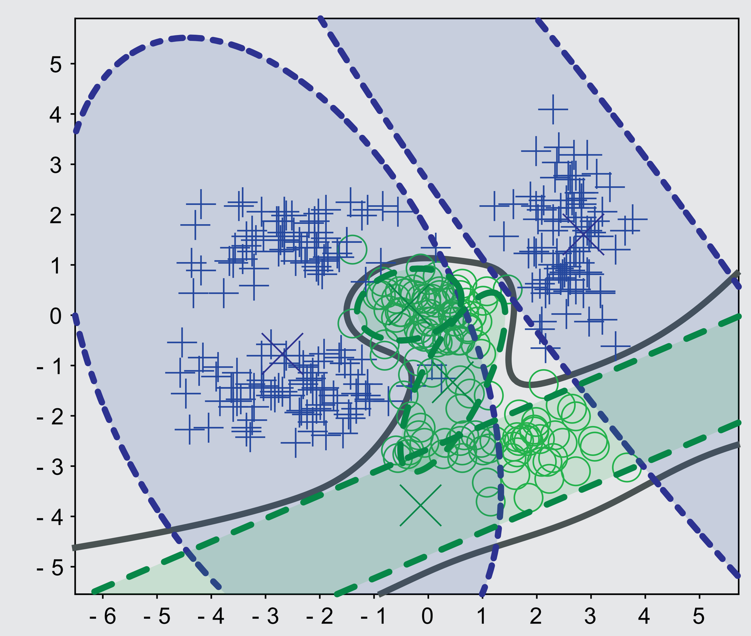


One Paradigm – Two Training Methods

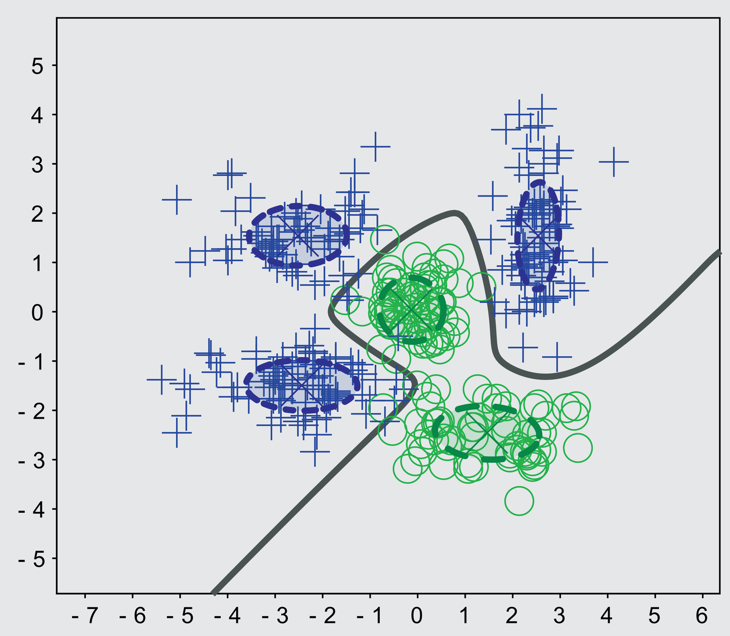
RBF can be trained in different ways:

- **Resilient Propagation** (RPROP)
  - **Discriminative** Training
  - Gradient-based approach (first-order)
  - Objective function: least-squares or cross-entroy error
  - Number of hidden neurons must be specified in advance
- **Variational Inference** (VI)
  - **Generative**Training
  - Bayesian approach, Maximum Likelihood parameter estimates (θ: set of all parameters)
  - Objective function: log-likelihood $\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\theta})$
  - Number of components found automatically

Visual Comparison of Trained Models



RBF trained with RPROP (**discriminative**)



RBF trained with VI (**generative**)

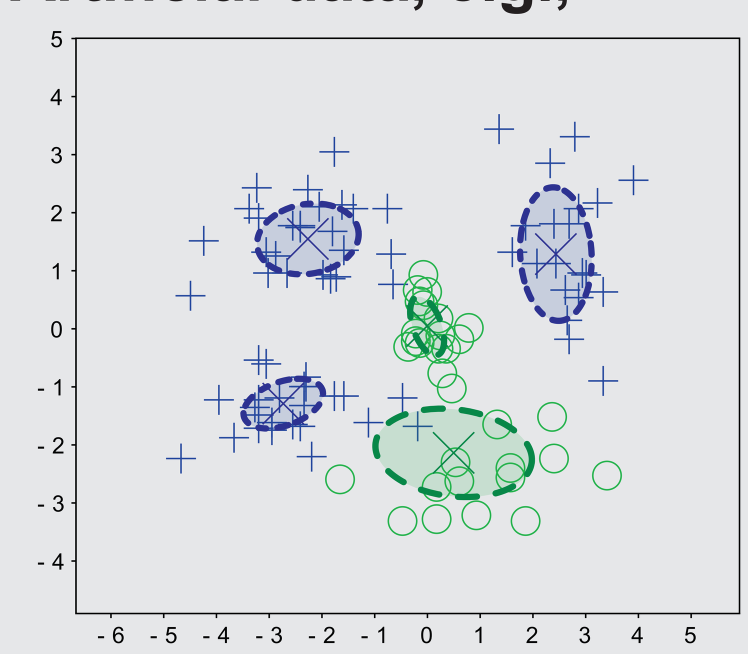
Numerical Measures for Comparison

- **Discriminative Properties:**
  Classification Error
- **Generative Properties:**
  Representativity based on an approximation of the Kullback-Leibler divergence of true model p and trained model q:

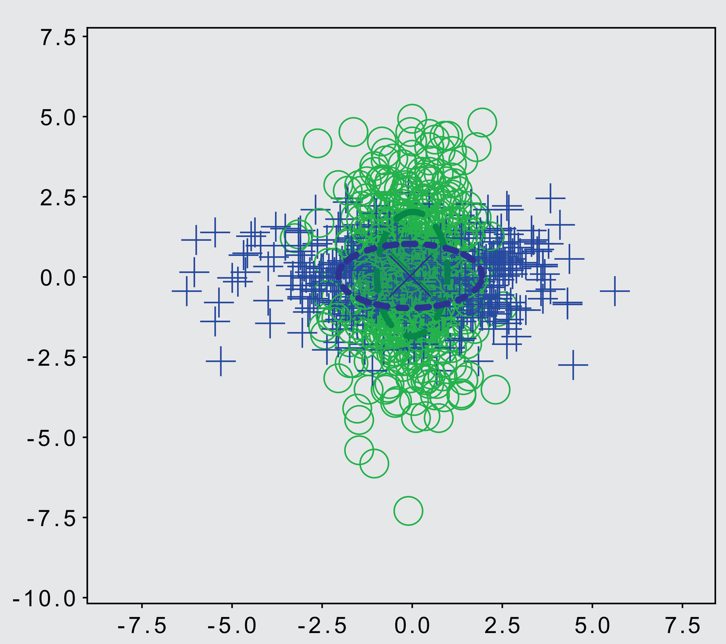
$$\widehat{\mathrm{KL}}_2(p,q) = \frac{1}{2N}\left(\sum_{\mathbf{x}_n} \ln \frac{p(\mathbf{x}_n)}{q(\mathbf{x}_n)} + \sum_{\mathbf{x}_n} \frac{q(\mathbf{x}_n)}{p(\mathbf{x}_n)} \ln \frac{q(\mathbf{x}_n)}{p(\mathbf{x}_n)}\right)$$

Datasets

- **Artificial data, e.g.,**



Sparse data



Overlapping classes

- **Real data from the Research Project ELENA**
  - Clouds
  - Satimage
  - Phoneme

Results

- RPROP reliably yields solutions with good discriminative properties
- RPROP does not support generative properties at all
- Training-times of RPROP are typically higher than those of VI
- VI reliably leads to solutions with good representativity if data is approximately normally distributed
- If distribution assumptions are met: Classification performance similar to RPROP
- Otherwise: Often worse than RPROP
- Choice between generative or discriminate classifier should depend on the particular application