Genetic Algorithm with Species for Regularization Network Metalearning

Petra Vidnerová

Department of Theoretical Computer Science
Institute of Computer Science
Academy of Sciences of the Czech Republic

ITAT 2010

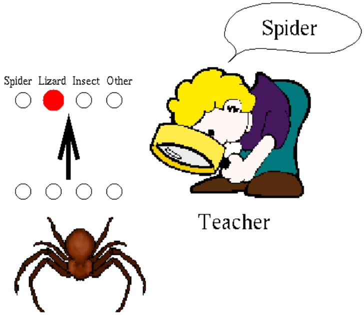
Outline

- Introduction - supervised learning
- Regularization networks
- Meta-parameters
- Genetic parameter search
- Experimental results
- Summary and future work

Supervised Learning

Learning

- given set of data samples
- find underlying trend, description of data



Supervised learning

- data – input-output patterns
- create model representing IO mapping
- classification, regression, prediction, etc.

Regularization Networks

Regularization Networks

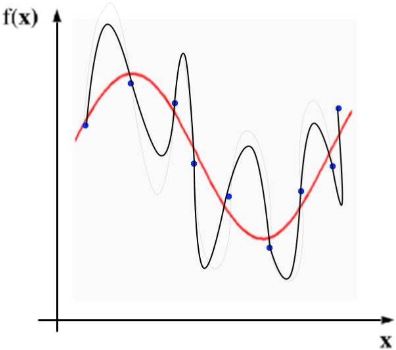
- method for supervised learning
- a family of feedforward neural networks with one hidden layer
- derived from regularization theory
- very good theoretical background

Our Focus

- we are interested in their real applicability
- setup of explicit parameters

Learning from Examples - Problem Statement

- **Given:** set of data samples $\{(\vec{x}_i, y_i) \in R^d \times R\}_{i=1}^N$
- **Our goal:** recover the unknown function or find the best estimate of it



Regularization Theory

Empirical Risk Minimization:

- find f that minimizes $H[f] = \sum_{i=1}^{N}(f(\vec{x}_i) - y_i)^2$
- generally ill-posed
- choose one solution according to a priori knowledge (*smoothness, etc.*)

Regularization approach

- add a **stabiliser** $H[f] = \sum_{i=1}^{N}(f(\vec{x}_i) - y_i)^2 + \gamma\Phi[f]$

Derivation of Regularization Network

Form of the Solution

- for a wide class of stabilizers the solution has a form

$$f(x) = \sum_{i=1}^{N} w_i G(\vec{x} - \vec{x}_i)$$

- where weights w_i satisfy

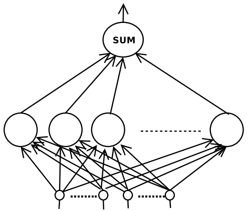
$$(\gamma I + G)\vec{w} = \vec{y}$$

- represented by feed-forward neural network with one hidden layer

Derivation of Regularization Network

Regularization Network

$$f(x) = \sum_{i=1}^{N} w_i G(\vec{x} - \vec{x}_i)$$



- function G called basis or kernel function
- choice of G represents our knowledge or assumption about the problem
- choice of G is crucial for the generalization performance of the network

RN learning algorithm

Basic Algorithm

1. set the centers of kernel functions to the data points
2. compute the output weigths by solving linear system

$$(\gamma I + K)\vec{w} = \vec{y}$$

Advantages and Disadvantages

- algorithm simple and efective
- choice of γ and kernel function is crucial for the performance of the algorithm (cross-validation)

Meta-parameters

Parameters of the Basic Algorithm

- kernel type
- kernel parameter(s) (i.e. width for Gaussian)
- regularization parameter γ

How we estimate these parameters?

- kernel type usually by user
- kernel parameter and regularization parameter by cross-validation
- in this work: all parameters by genetic approach

Role of Kernel Function

Choice of Kernel Function

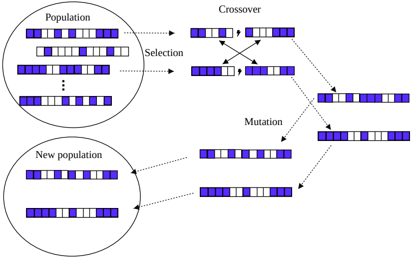
- choice of a stabilizer
- choice of a function space for learning (hypothesis space)

Role of Kernel Function

- represent our prior knowledge about the problem
- *no free lunch* in kernel function choice
- should be chosen acording to the given problem
- what functions are good first choice?

Genetic Algorithm

- stochastic optimisation technique
- work with population of possible solutions
- operators selection, crossover, mutation



Genetic Algorithm Search

Individuals

- individuals coding RN meta-parameters $I = \{K, p, \gamma\}$

Individual used for search including kernel type:

type of kernel	kernel parameters	reg. parameter

Individual used for Gaussian kernels:

width	reg. parameter

Co-evolution

- subpopulations corresponding to different kernel functions
- selection on the whole population
- crossover on subpopulations

b_1	gamma_1
b_2	gamma_2

crossover $\Rightarrow$

a_1*b_1+(1-a_1)*b_2)	c_1*gamma_1+(1-c_1)*gamma_2
a_2*b_1+(1-a_2)*b_2)	c_2*gamma_1+(1-c_2)*gamma_2

Experiments

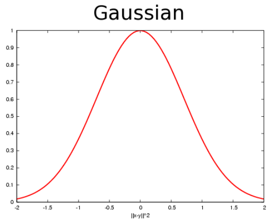
Data

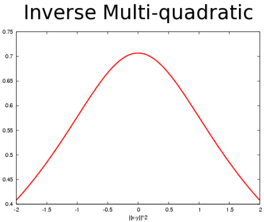
- benchmark data sets - Proben1 data repository

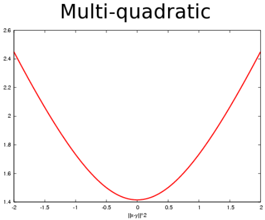
Methodology

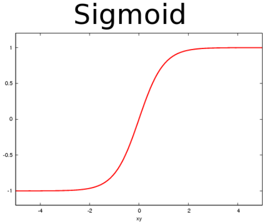
- separate data for training and testing
- find suitable kernel function and γ on training set by genetic parameter search
- learn on training set (estimation of weights w)
- evaluate error on testing set - generalization ability

Kernel Functions

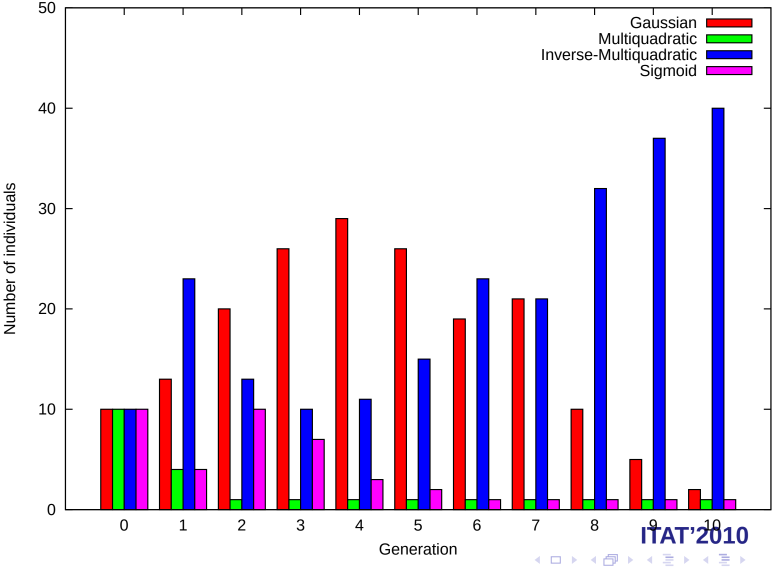




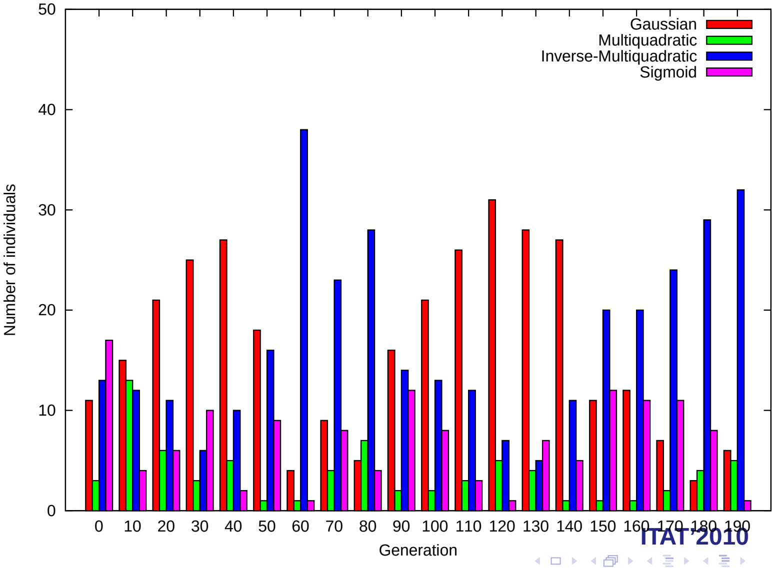




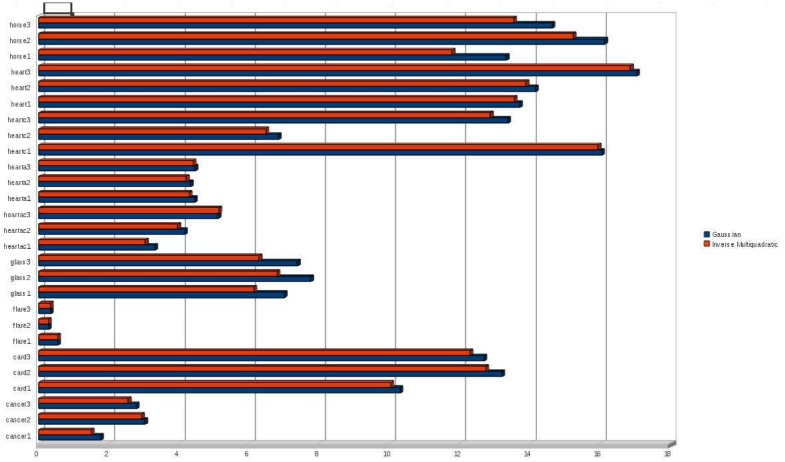
Subpopulations during Evolution - Tournament Selection



Subpopulations during Evolution - Roulette-wheel Selection



Comparison with Gaussian Kernel



Summary and future work

Summary

- learning with RN networks described
- role of kernel function discussed
- genetic parameters search
- best kernel inverse-multiquadratic

Work in progress and future work

- kernel functions for other data types (categorical data, etc.)
- composite types of kernels

Thank you! Questions?