An Overview of Nonlinear Black-box Modelling

Elia Liitiäinen
eliitiai@cc.hut.fi

Contents

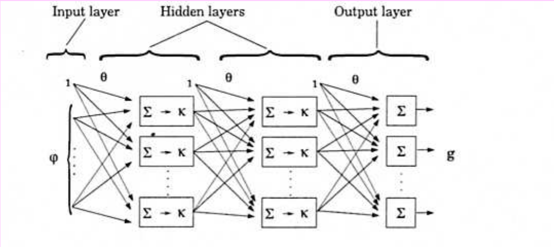
Introduction

- There is many possible model structures for nonlinear black-box modelling
- In this presentation a short overview is given

Neural Networks

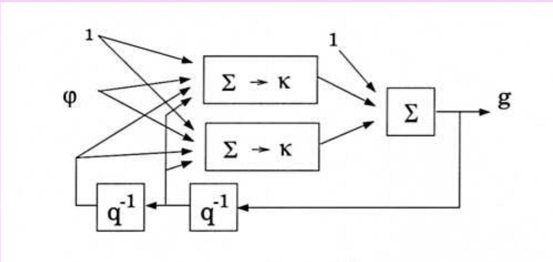
- Neural networks are a general class of approximators
- Suppose $\phi_k(\cdot;\theta_k) : R^d \to R$ are (nonlinear) mappings
- A two layer (one dimensional) neural network is a function of the form $g(t; w, \theta) = \sum_{k=1}^{l} w_k \phi_k(t; \theta_k)$.
- w and θ are the parameters of the network
- Each activation function ϕ_k corresponds to one neuron

Multilayer Neural Networks



- By adding multiple layers the approximation properties can be improved
- Parameter estimation becomes more difficult
- Theoretically the MLP is a difficult topic as approximation results are hard to obtain
  - http://math.tkk.fi/ ggripenb/ggpub.htm

Recurrent Neural Networks



- Time dependent networks

Choice of the Nonlinear Mapping (1)

- We suppose that each neuron is a mapping of the form

$$\phi_k(t) = \kappa(t, \beta_k, \gamma_k)$$

- Next some possibilities for this mapping will be listed

Choice of the Nonlinear Mapping (2)

- Splines
- Radial functions

$$\phi_k(t) = \kappa(\|t - \gamma_k\|_k)$$

- Ridge functions

$$\phi_k(t) = \kappa(\beta_k^T t + \sigma_k)$$

  - A typical choice for κ is tanh
  - Any continuous functions can be approximated with this nonlinearity
- Wavelets

Choosing parameters

- Parameters can be chosen by minimizing some cost function
- A common choice is the maximum likelihood cost:

$$\sum_{i=1}^{N}(y_i - g(x_i;\theta))^2$$

- The pairs (x_i, y_i) form the training set
- A nonlinear optimization technique like the conjugate gradient method can be used
- To avoid overfitting a test set can be used

Basis Function Networks

- First a basis $\mathcal{G}$ in some function space is chosen
- Let $(\phi_k)_{k=1}^N$ be a finite subset of $\mathcal{G}$
- A basis function network is a neural network of the form

$$g(t; w) = \sum_{k=1}^{N} w_k \phi_k(t)$$

- The approximator is linear in parameters which simplifies parameter estimation

Choice of the Basis (1)

- The most common choice is a radial basis:

$$\phi_k(t) = \kappa(||(t - \gamma_k)||_k)$$

- The centers for the basis can be chosen with vector quantization
- Another possibility is using wavelets which are functions of the form

$$\phi_k(t) = \kappa(\beta_k(t - \gamma_k))$$

Choice of the Basis (2)

- The set $(\phi_k)_{k=1}^{N}$ can be chosen in many ways
- Residual based selection
  - At each step those basis functions that best fit the residual are chosen
- Stepwise selection by orthogonalization
- Backward elimination
  - At each step functions are removed so that the error grows as little as possible

Conclusions

- The most commonly used approximators are the MLP (ridge functions) and radial basis function networks (RBFN)
- For RBFNs simple and fast techniques can be used for basis selection
- Ridge functions have better approximation properties but lead to a nonlinear optimization problem