LUND UNIVERSITY

# Introduction to Neural Networks

Jacek Malec
Department of Computer Science
Lund University
email: jacek.malec@cs.lth.se

19th March 2009

AI@CS

Department of Computing Science



LUND UNIVERSITY

## Introduction to Neural Networks

Topics for today:

1. what are neural networks?
2. what is their hypothesis space?
3. what is backpropagation algorithm?
4. what are recurrent neural networks?

AI@CS

Department of Computing Science



LUND UNIVERSITY

## Connectionist Models

- Neuron switching time 0.001 second
- Number of neurons $10^{10}$
- Connections per neuron $10^{4-5}$
- Scene recognition time 0.1 second
- 100 inference steps doesn't seem like enough

$\rightarrow$ much parallel computation

AI@CS

Department of Computing Science



LUND UNIVERSITY

## When to Consider Neural Networks

- Input is high-dimensional discrete or real-valued (e.g. raw sensor input)
- Output is discrete or real valued
- Output is a vector of values
- Possibly noisy data
- Form of target function is unknown
- Human readability of result is unimportant

AI@CS

Department of Computing Science

## More on backpropagation

- Gradient descent over entire *network* weight vector
- Easily generalized to arbitrary directed graphs
- Will find a local, not necessarily global error minimum
  - In practice, often works well (can run multiple times)
- Often include weight *momentum* $\alpha$

$$\Delta w_{i,j}(n) = \eta \delta_j x_{i,j} + \alpha \Delta w_{i,j}(n-1)$$

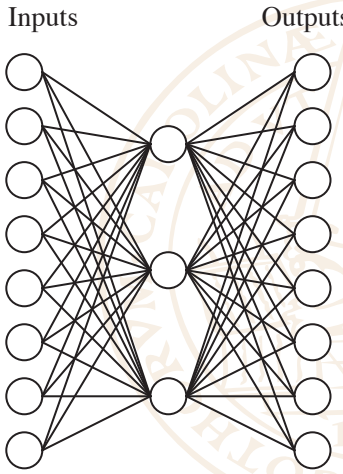


## More on backpropagation

- Minimizes error over *training* examples
  - Will it generalize well to subsequent examples
- Training can take thousands of iterations → slow!
- Using network after training is very fast



## Learning hidden layer representations






## Learning hidden layer representations

A target function:

| Input | | Output |
|---|---|---|
| 10000000 | → | 10000000 |
| 01000000 | → | 01000000 |
| 00100000 | → | 00100000 |
| 00010000 | → | 00010000 |
| 00001000 | → | 00001000 |
| 00000100 | → | 00000100 |
| 00000010 | → | 00000010 |
| 00000001 | → | 00000001 |

Can this be learned?

## Learning hidden layer representations

LUND UNIVERSITY

Learned hidden layer representation:

| Input | | Hidden Values | | | | Output |
|---|---|---|---|---|---|---|
| 10000000 | → | .89 | .04 | .08 | → | 10000000 |
| 01000000 | → | .01 | .11 | .88 | → | 01000000 |
| 00100000 | → | .01 | .97 | .27 | → | 00100000 |
| 00010000 | → | .99 | .97 | .71 | → | 00010000 |
| 00001000 | → | .03 | .05 | .02 | → | 00001000 |
| 00000100 | → | .22 | .99 | .99 | → | 00000100 |
| 00000010 | → | .80 | .01 | .98 | → | 00000010 |
| 00000001 | → | .60 | .94 | .01 | → | 00000001 |

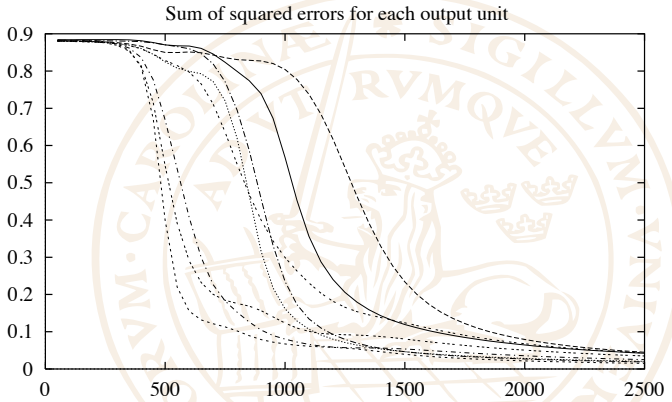
AI@CS

Department of Computing Science



## Learning hidden layer representations

LUND UNIVERSITY



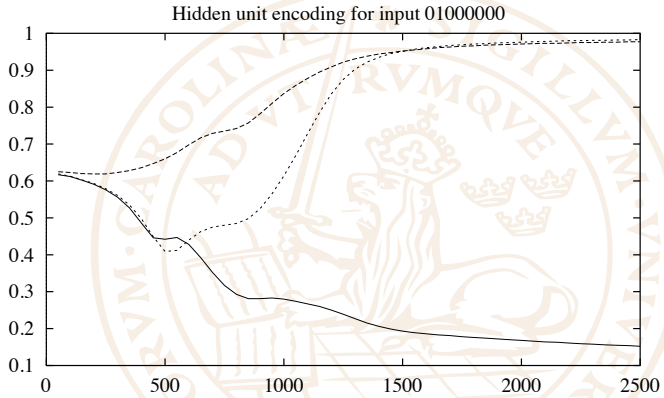

AI@CS

Department of Computing Science



## Learning hidden layer representations

LUND UNIVERSITY



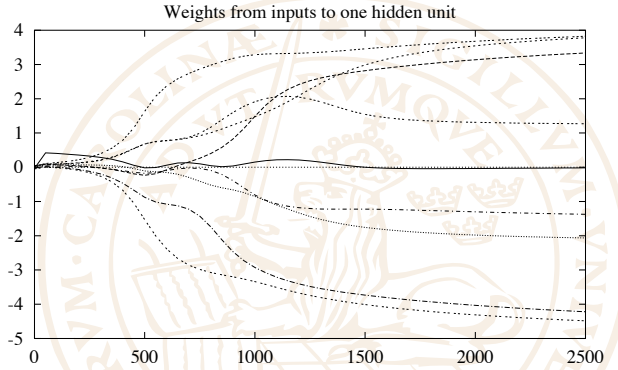

AI@CS

Department of Computing Science



## Learning hidden layer representations

LUND UNIVERSITY




AI@CS

Department of Computing Science

## Convergence of backpropagation

Gradient descent to some local minimum

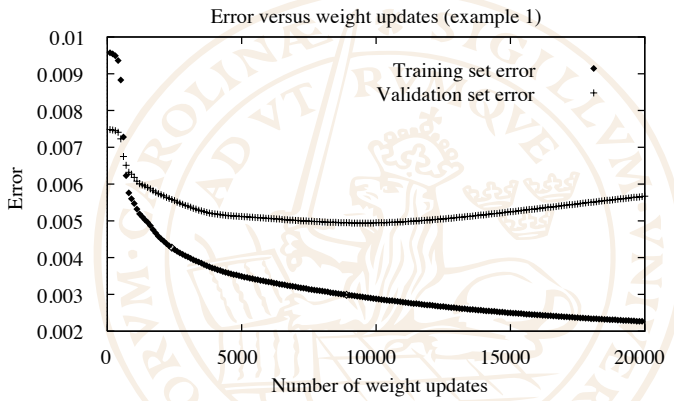
- Perhaps not global minimum...
- Add momentum
- Stochastic gradient descent
- Train multiple nets with different inital weights

Nature of convergence

- Initialize weights near zero
- Therefore, initial networks near-linear
- Increasingly non-linear functions possible as training progresses

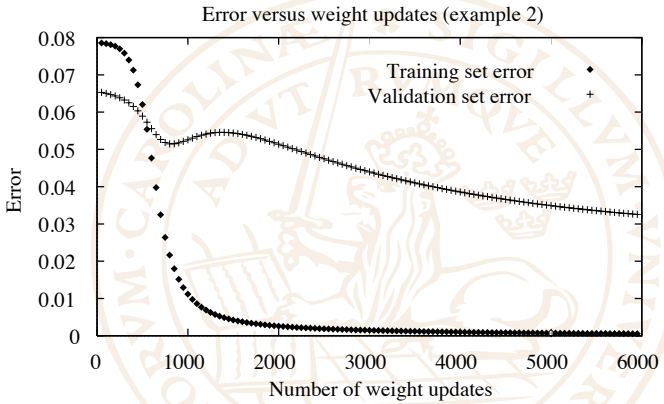


## Overfitting






## Overfitting





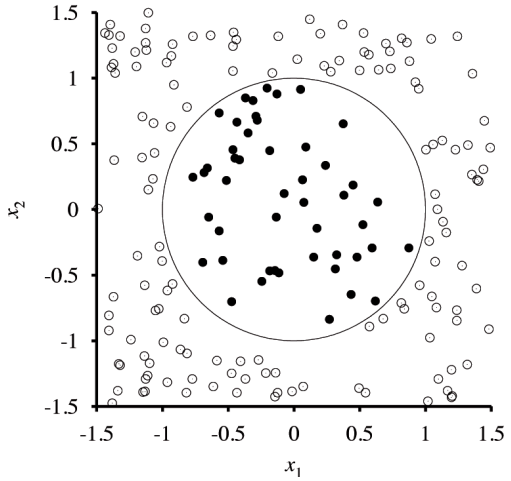

## Kernel methods

Support Vector Machines (SVM)
Kernel Machines

efficient learning AND expressiveness

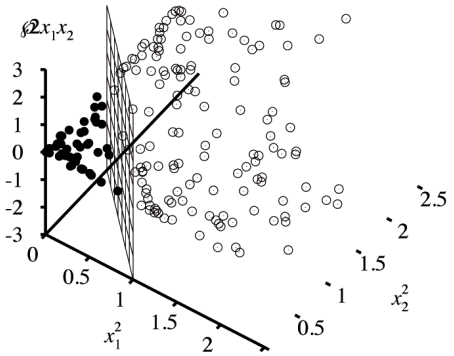
unsupervised kernel parameter fitting
+
supervised weight fitting

## Kernel machines



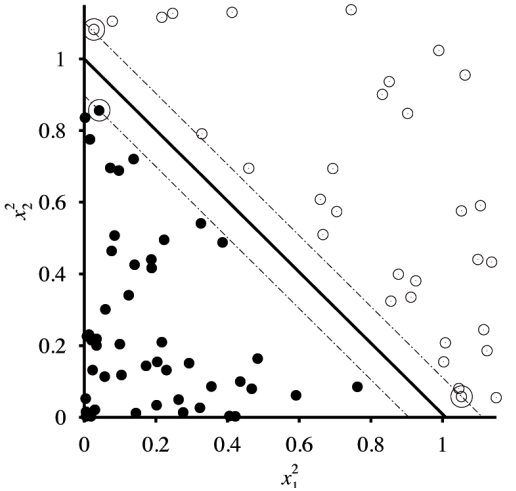



## Kernel machines






## Kernel machines






## Recurrent networks