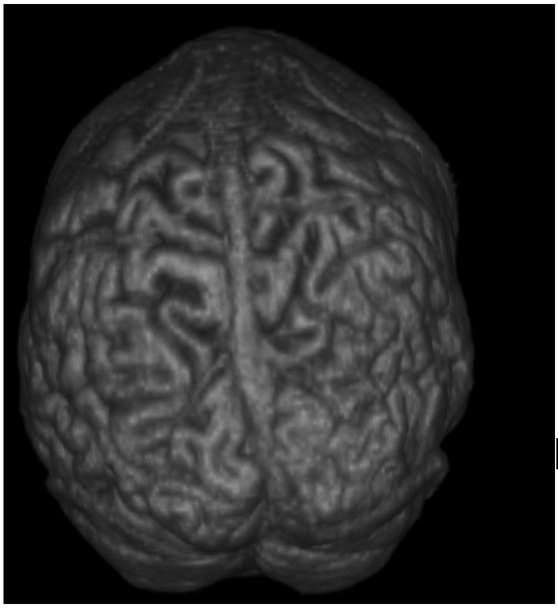


EGR 301

Artificial Neural Networks

Prof. Glenn Ellis

Spring 2005

Objectives

1. Ability to use a backpropagation, feed-forward ANN.
2. Acquire some insight into how they work, their limitations, etc.

How do we teach a child to differentiate cats from dogs?


Expert Systems

Teach rules

Cats say meow. Dogs say woof.

Examples

Medicine
Water treatment

ANNs

1. Train
Show example
Compare child's and actual answer
Reward/Correct

Iterate

Iterate

2. Test
Show new pictures

3. Apply
Interact with cats and dogs

ANNs

1. Train
Show example
Compare child's and actual answer
Reward/Correct

Iterate

1a. Validate
Pre-test to see if we should stop training.

2. Test
Show new pictures

Iterate

3. Apply
Interact with cats and dogs

Note

Need training, validation, test sets.

Ann as good as data set.

Ann learns relationships.

What can go wrong?

Bad ANN

Error in dataset

Not enough data

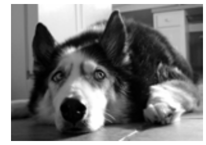
Not enough independent data

Not random sample

Apply outside domain

What can go wrong?

Bad ANN

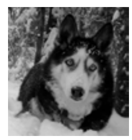


Error in dataset

Not enough data

Not enough independent data

Not random sample



Apply outside domain

What can go wrong?

Bad ANN

Error in dataset

Not enough data

Not enough independent data

Not random sample

Apply outside domain

ANNs solve some classical AI problems

Pattern recognition

100 step constraint

Graceful degradation

Multiple soft constraints

Knowledge relevance

Credit Card Application

How do we create an expert system?

Credit Card Application

Expert System – Interview experts and decide on rules. Apply rules.

Credit Card Application

How do we create an ANN?

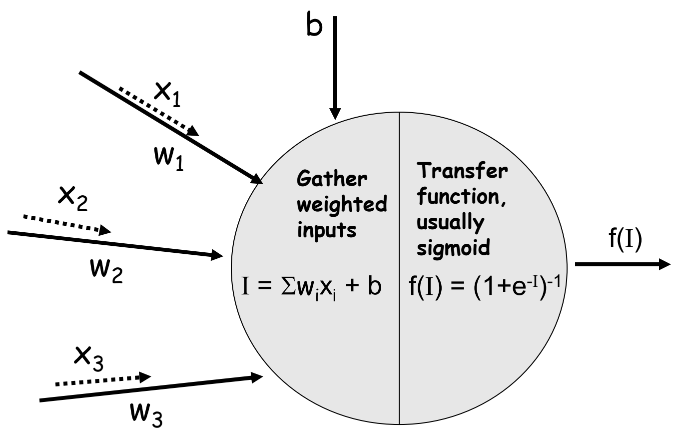
Credit Card Application

1. Get data.
2. ANNs – Train, test and apply.

Credit Card Application

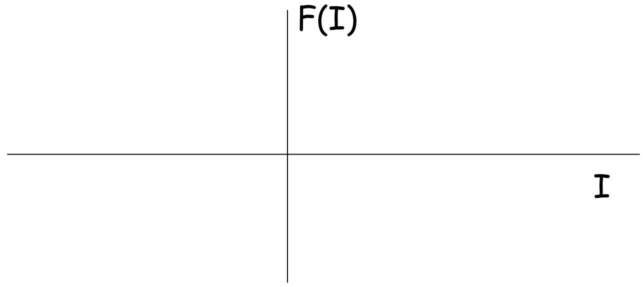
Any ethical concerns?

<u>Neuron:</u> Gathers signals from synapses, processes, sends output



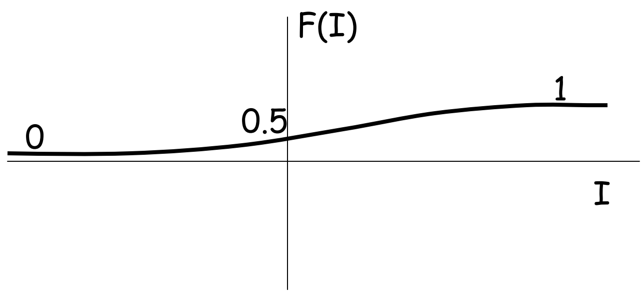
What does sigmoid function look like?

$f(I) = (1+e^{-I})^{-1}$



What does sigmoid function look like?

$f(I) = (1+e^{-I})^{-1}$



Create an ANN to check credit.

Inputs: ???

Outputs: ???

I

Notes on hidden layer

May have many layers.

Allows deeper (non-linear) learning.

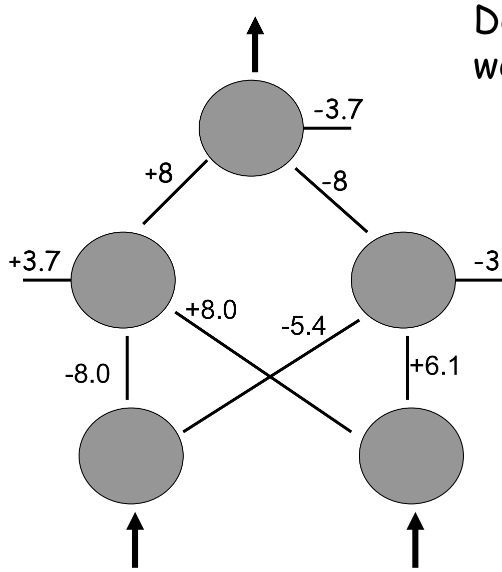
Sees weighted inputs.

Get # of layers and neurons by trial and error, genetic algorithms, etc..

I

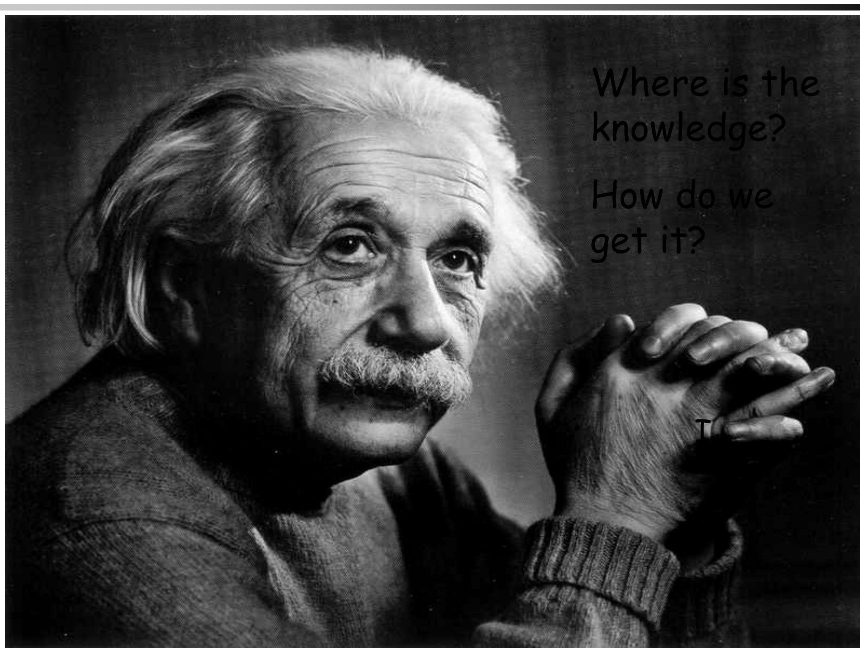
ROT for starting: # hidden neurons = (# inputs + # outputs) / 2.

Do these weights work?



Not XOR

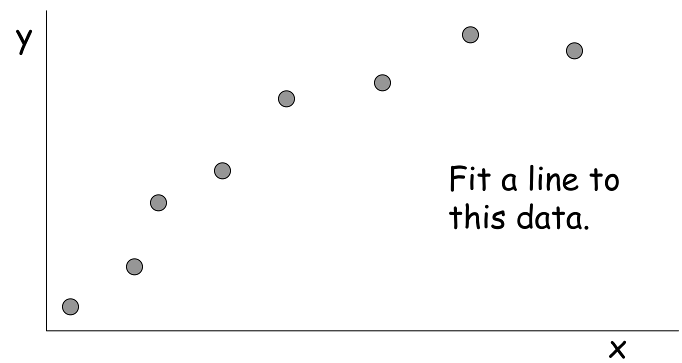
0,0 → 1
1,1 → 1
0,1 → 0
1,0 → 0



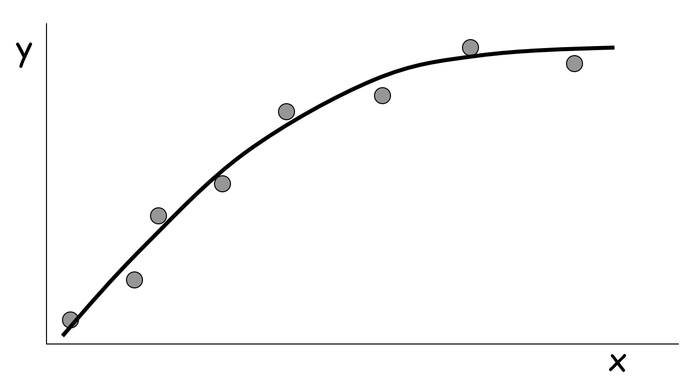
Supervised Training

1. Show ANN inputs
2. Compute output(s)
3. Compute error, $\Sigma(\text{output} - \text{target})^2$
4. Is error small enough? If yes, stop.
5. No, adjust weights (using backpropagation) and go back to (1).

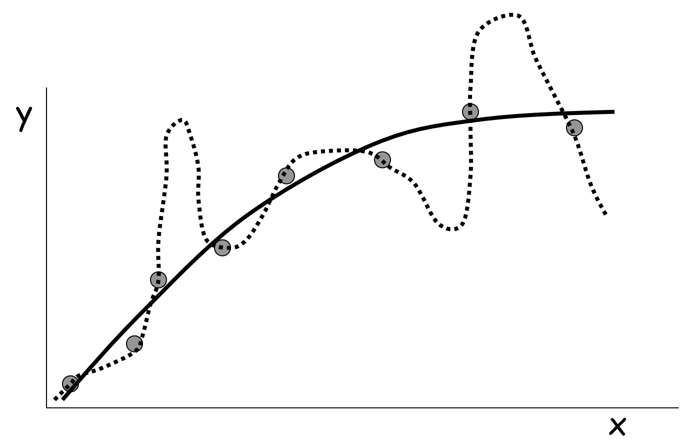
How do we know it has learned something?



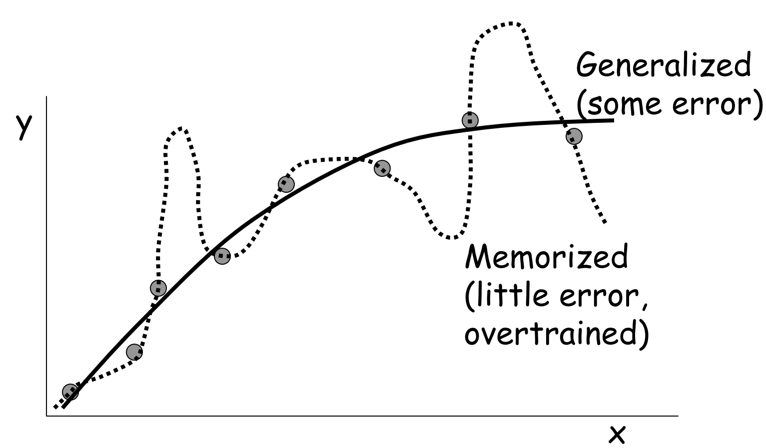
Human attempt



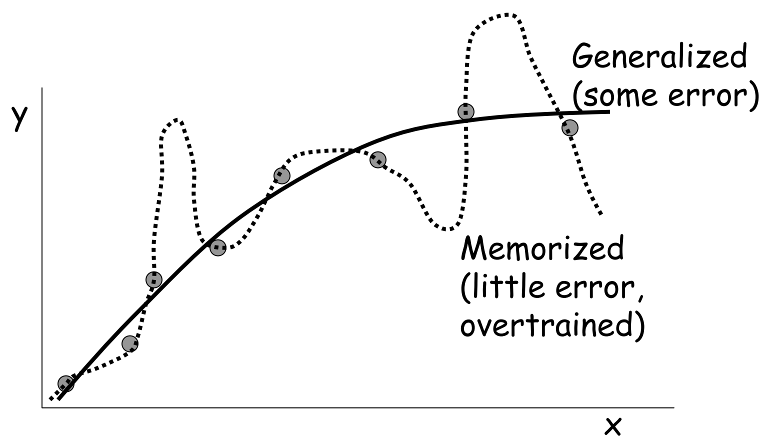
ANN after lots of training



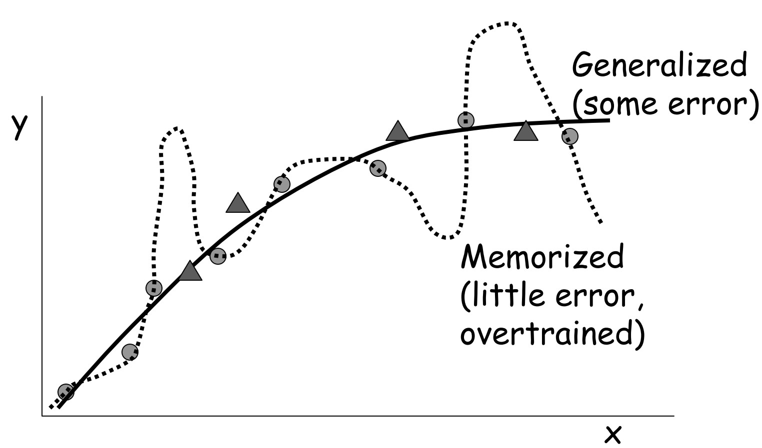
What does this mean?



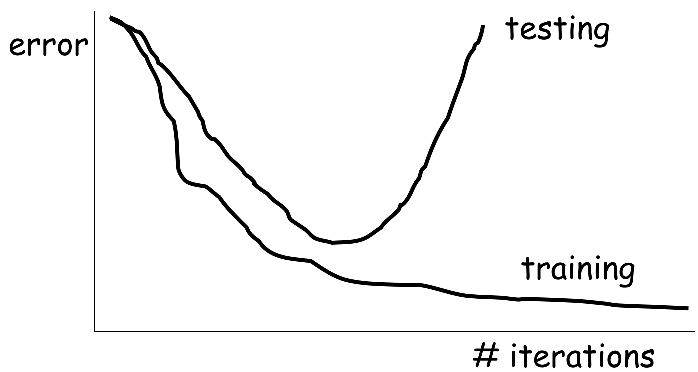
How do we know when to stop if this graph is in 130 dimensions?



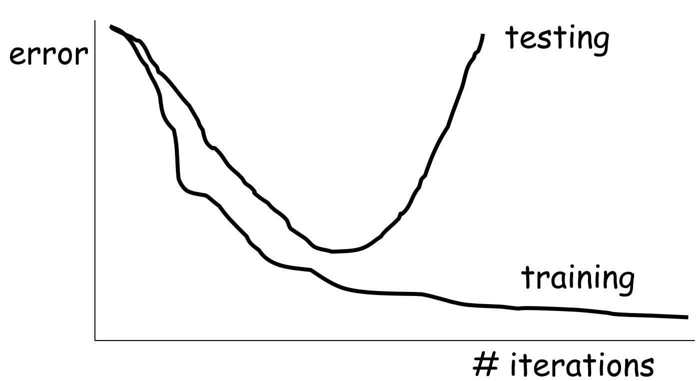
Test it on data it hasn't seen.



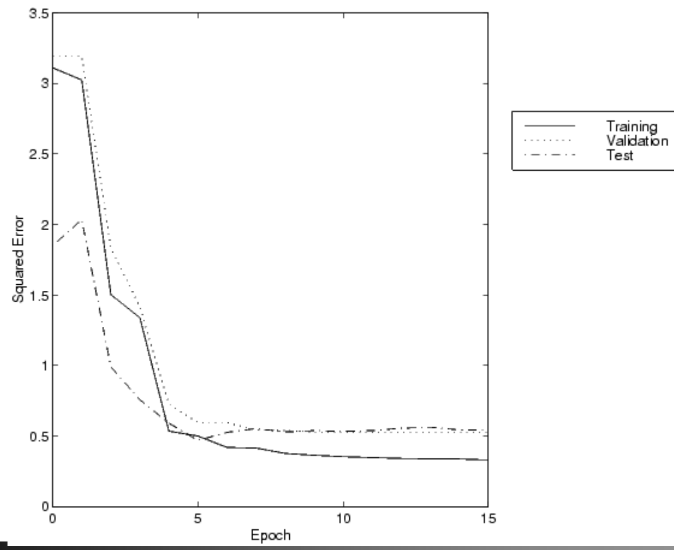
Early Stopping



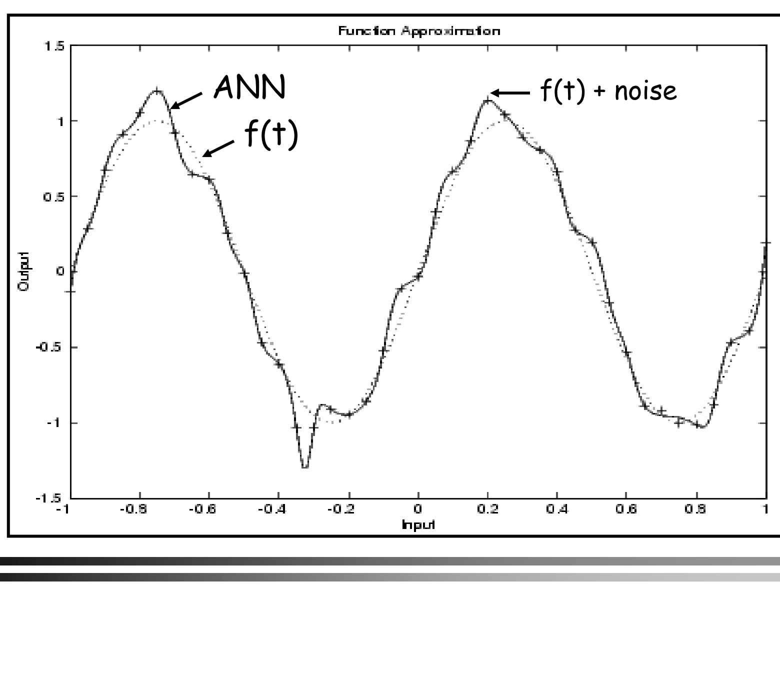
But this is sort of cheating, how?



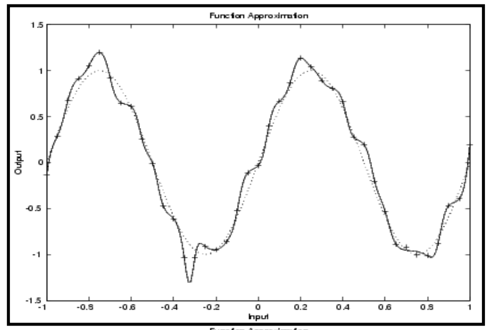
Is there over-training in this example?



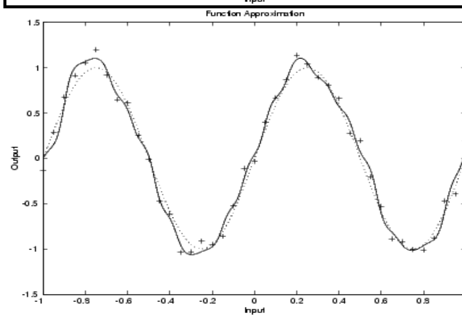
MATLAB example with overtraining



Overtrained



With early
stopping



What if it doesn't do well in testing?

1. Overtrained
2. No underlying relationship
   Potsdam Water Treatment,
   Stamford Wastewater Treatment Plant
3. ANN can't learn it.
4. Insufficient data.

What if it doesn't do well in testing?

1. Overtrained
2. No underlying relationship
   Potsdam Water Treatment,
   Stamford Wastewater Treatment Plant
3. ANN can't learn it.
4. Insufficient data.

Relate to dog/cat.

Cure.

With limited data, how much should be used for training and testing?

Answer: depends

What does putting more data into the <u>training</u> set get us?

What does putting more data into the <u>testing</u> set get us?

With limited data, how much should be used for training and testing?

Answer: depends

What does putting more data into the <u>training</u> set get us? Higher chance that it learns.

What does putting more data into the <u>testing</u> set get us? Higher confidence that it has learned.

ROT

10 – 20 independent data points for each i/o neuron.

ROT

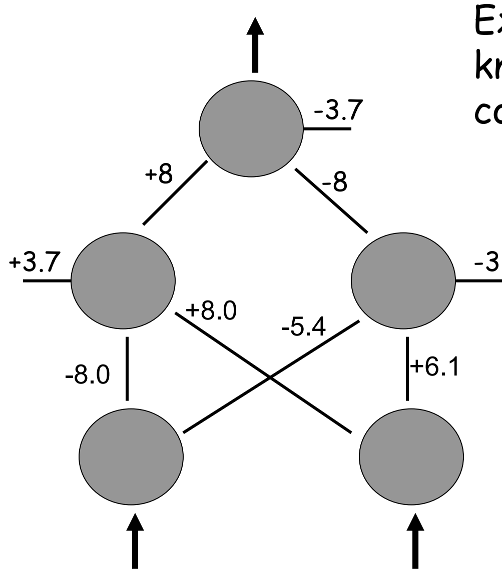
10 – 20 **independent** data points for each i/o neuron.

ROT

10 – 20 **independent** data points for each i/o neuron.

90% of data in training set

10% of data in testing set



Explain the knowledge it contains.

Not XOR

0,0 → 1
1,1 → 1
0,1 → 0
1,0 → 0

What can we do?

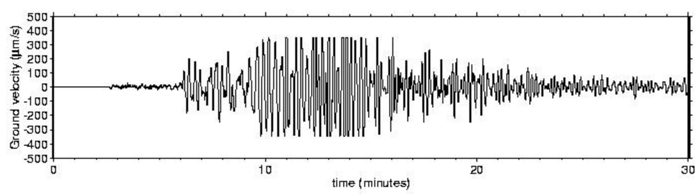
Vary one variable at a time and see how the output changes.

Main Applications

1. Pattern recognition

Train by looking at many patterns

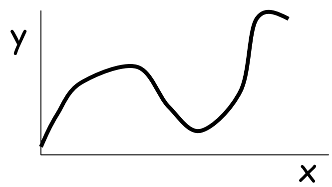
Examples: writing, speech, objects, seismograms



2. Function estimation

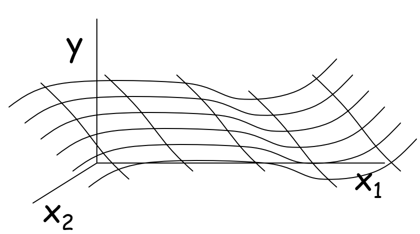
Y = f(x)

2. Function estimation



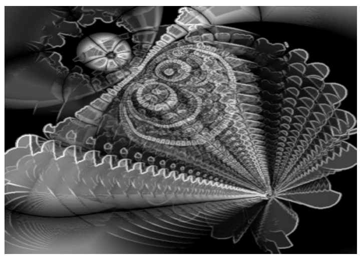
$Y = f(x)$

2. Function estimation



$Y = f(x_1,x_2)$

2. Function estimation

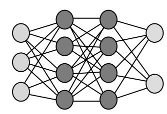


$$Y_{1-100} = f(x_{1-100})$$

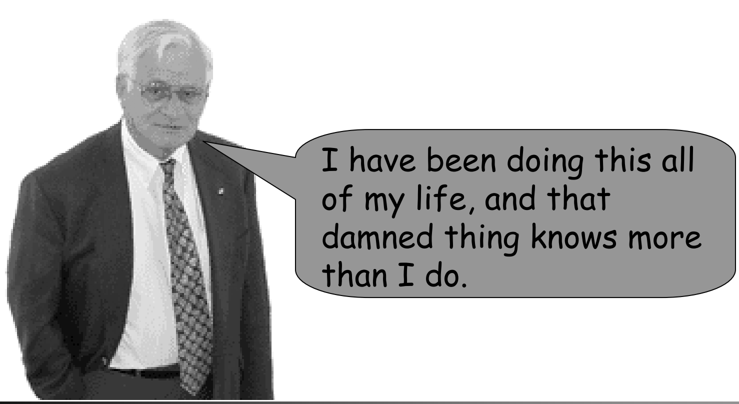
Example: Fiber-reinforced concrete beams

Example: Fiber-reinforced concrete beams

13 variables

(dimensions, loading, material variables) →  → Strength

Most accurate method in world 10 years ago.

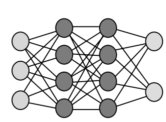


Most accurate method in world 10 years ago.

But, they'll never use it.

Boarding School Admissions

Geography
Grade
Sex
Minority
SSAT scores
Interview scores
Legacy

→  →

Admit
Waitlist
Reject


Boarding School Admissions

Results

Highly accurate

Most important factor?

- Geography
- Grade
- Sex
- Minority
- SSAT scores
- Interview scores
- Legacy


Boarding School Admissions

Results

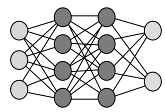
Highly accurate

Most important factor?

- Geography
- Grade
- Sex
- Minority
- SSAT scores
- Interview scores
- → Legacy

Ozone Water Disinfection

Dosing

Environmental conditions ➔  ➔ Virus conc.

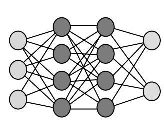
Results

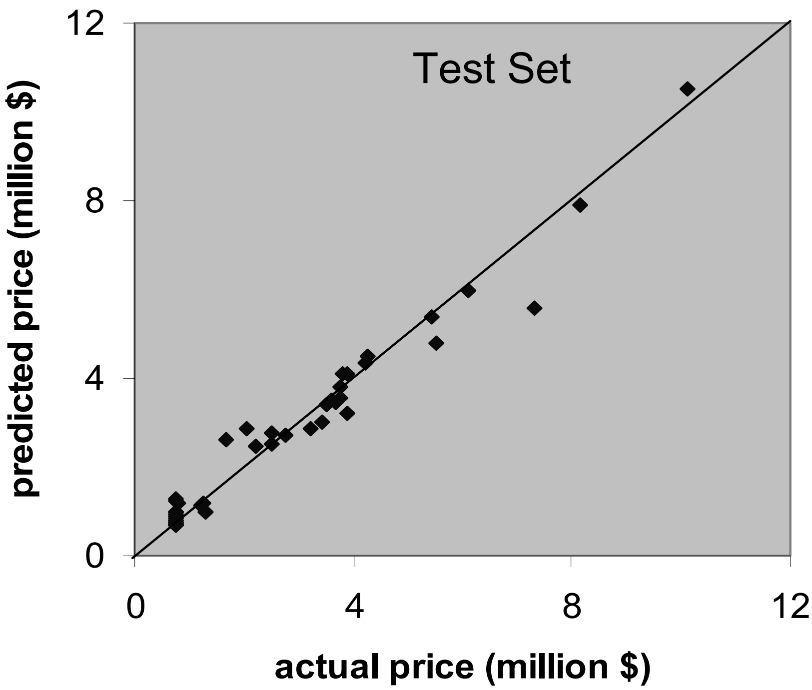
More efficient than EPA techniques

Published in: Environmental Engineering Science
Florida AI International Conference

Real Estate

Size (square ft., #bathrooms, #bedrooms, #garages)
Style (3 styles)
Land (acres, pool, courts, lakefront, oceanfront)
Location (9 neighborhoods)

➔  ➔ Price ($)



Applications

Detect price trends

Isolate variables
(value of saltwater frontage?)

Relate to secondary markets

Predict home improvement value

Appraisals

Back Propagation Training

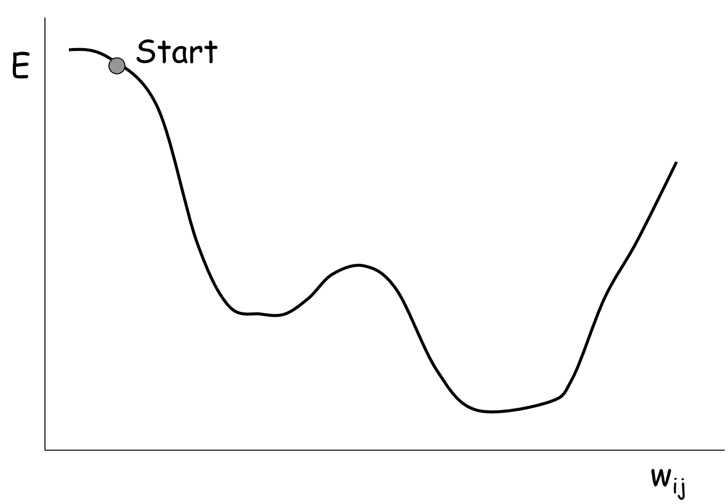
$$\Delta w_{ij} = - k \frac{\partial E}{\partial w_{ij}}$$

Change in error with respect to weight.

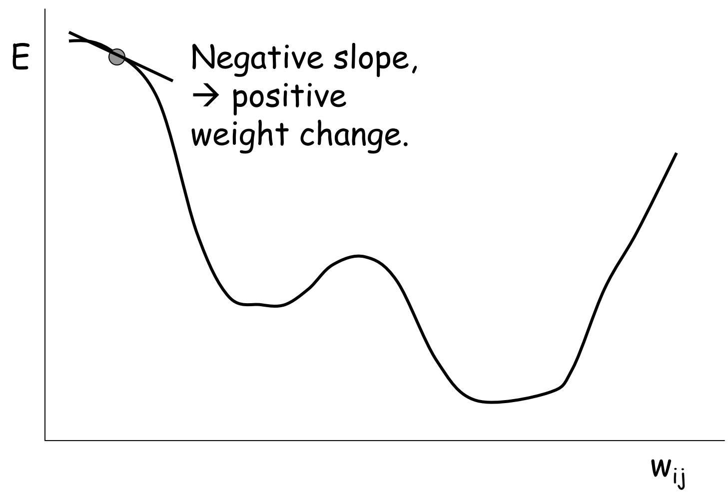
Go in direction to minimize error.

Learning rate

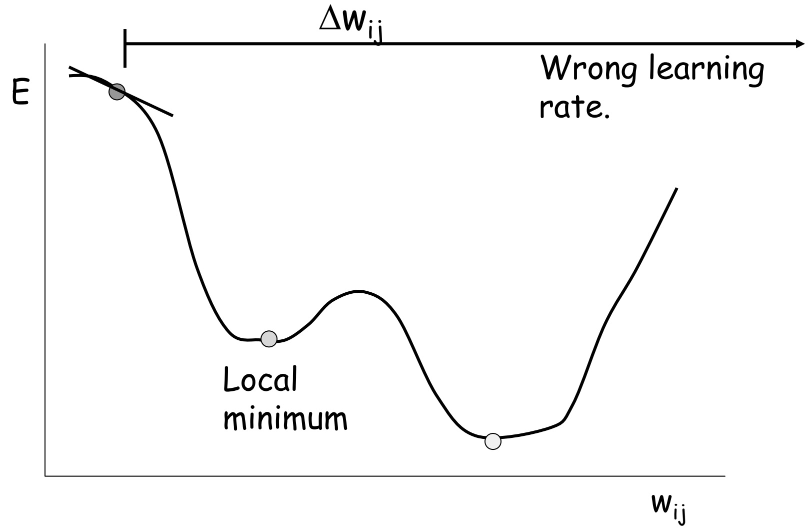
If we start here, which way will the weight change? Where do we want to go? What problems may occur?



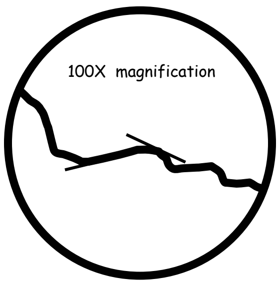
If we start here, which way will the weight change?
Where do we want to go? What problems may occur?



If we start here, which way will the weight change?
Where do we want to go? What problems may occur?



If we start here, which way will the weight change?
Where do we want to go? What problems may occur?



$$\Delta w_{ij}(n) = -k\frac{\partial E}{\partial w_{ij}} + \alpha \, \Delta w_{ij}(n-1)$$

where $0 < \alpha < 1$

Advice

1. Start with a low learning rate.
2. More complicated architectures need lower learning rates.
3. Need momentum to get out of oscillations.
4. Over specified networks will get confused.