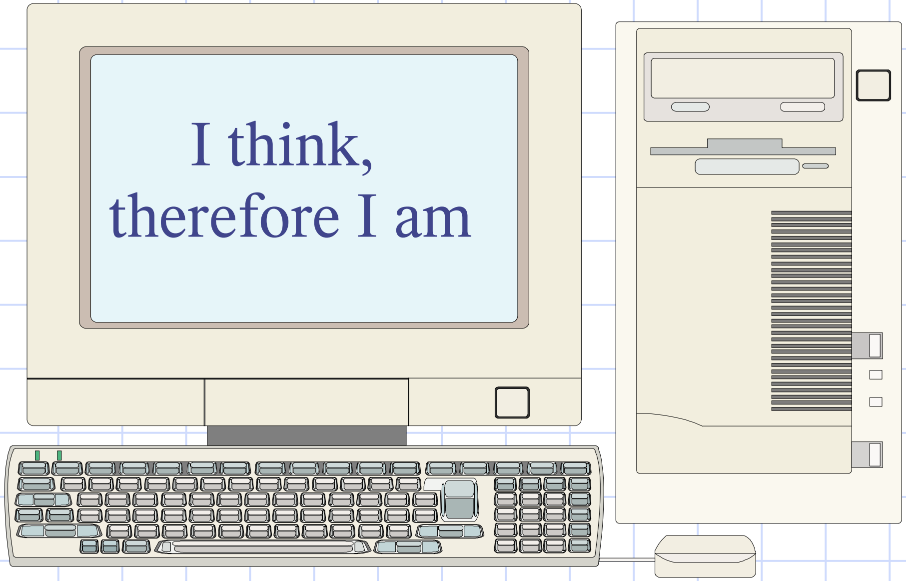
Neural Networks

A Statistical View

Brad Morantz PhD

The Future

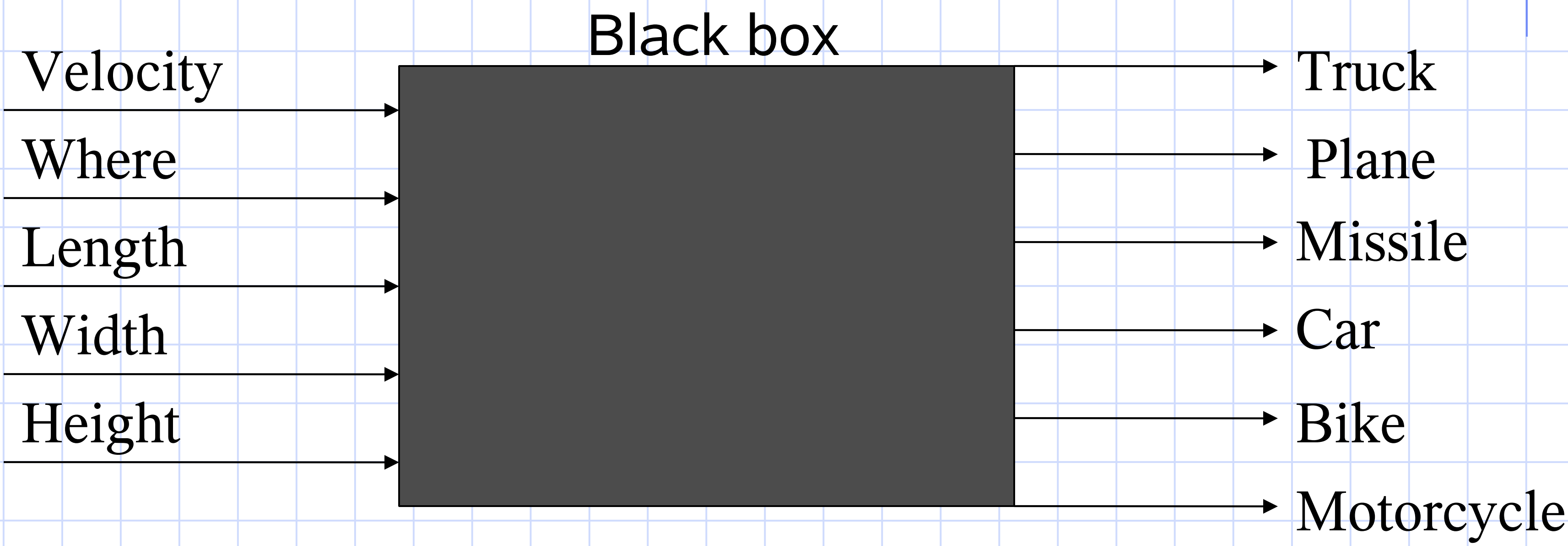


OK, so Descartes beat me to it, but this is in a different realm

Classification Problem

- Our sensors report:
  - Velocity - fuzzy: low, medium, or high
  - Sky or ground – categorical variable
  - Length – ratio variable
  - Width – ratio variable
  - height – ratio variable

How Do We Classify These?



Creating an Optimal Protein

- Causal model is not understood
- Solution: use an artificial neural network (ANN) with a genetic algorithm (GA)
  - Train ANN on known proteins
  - Use trained ANN as fitness function in GA
  - Use GA for exploited search for near optimal protein

Other Applications

- Image processing
  - Pixel: foreground or background classification
  - Non-linear filtering
  - Classification
  - Pattern recognition
- Radar
  - Tracker
  - Pattern recognition
- Medical
  - Diagnosis
  - Classification
  - Pattern recognition

More Applications

- Economic
  - Credit vetting
  - Forecasting
  - Fraud detection
- Military
  - Automatic target recognition
  - Steganography
  - Image processing
- The list goes on

Contents

What is a Neural Network?

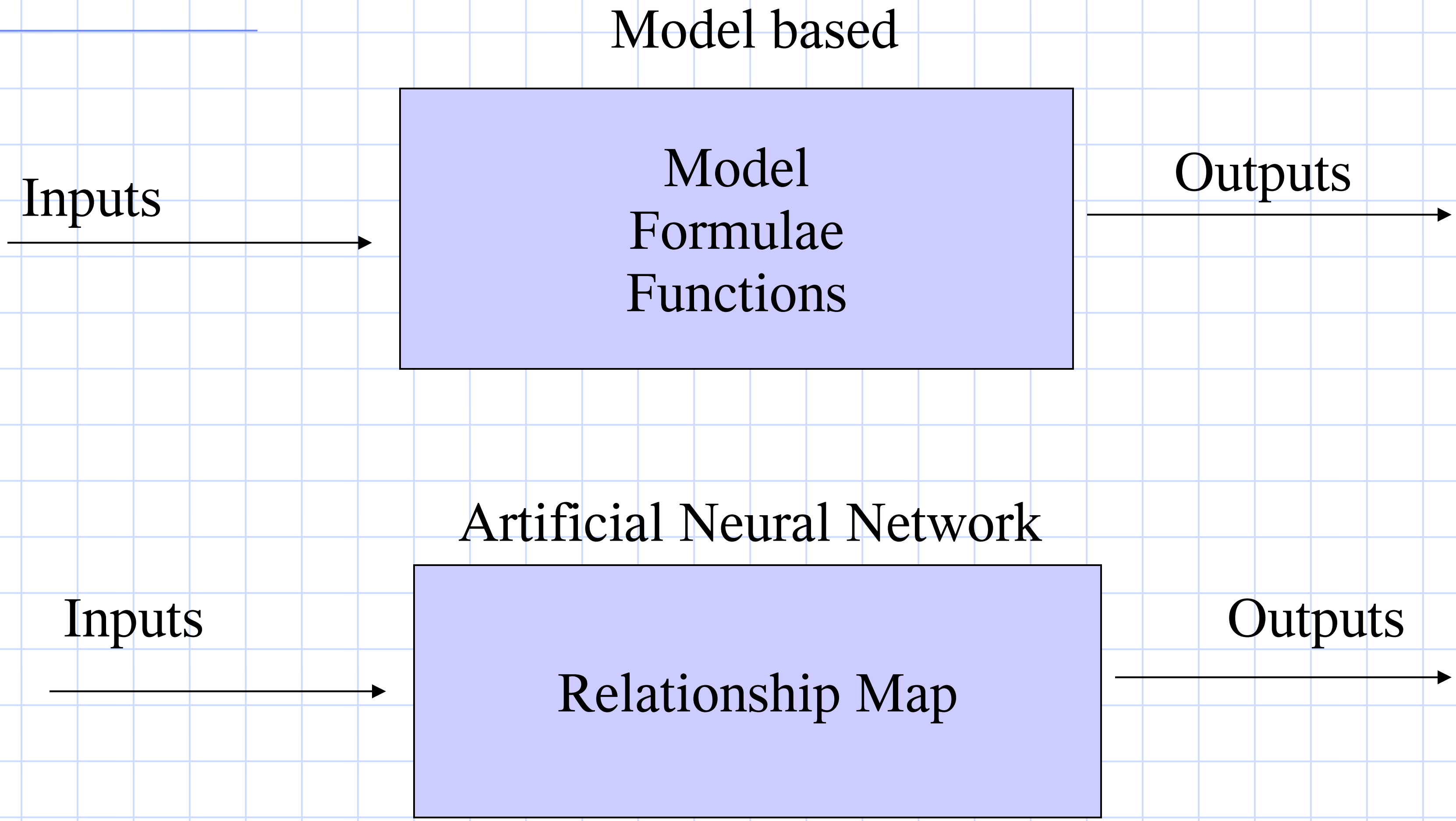
- A human Brain
- A porpoise brain
- The brain in a living creature
- A computer program
  - Emulates biological brain
  - Limited connections
- Specialized computer chip

What is an ANN?

(Artificial Neural Network)

- General function approximator
  - Imitates performance of original
  - Does not duplicate model
  - Does provide near or approximate results
  - It maps input to output
- Data driven
  - Does not understand causal model
  - Learns input to output relationship
  - Learns from supplied training data

Models



Model based
Inputs
Model
Formulae
Functions
Outputs
Artificial Neural Network
Inputs
Relationship Map
Outputs

What Can an ANN Use to Make Connections/Mapping?

- Learned Information
- From experience
- From historical data
- By example
- By organization

Four types of Functions

- Prediction and Time Series Forecasting
  - Like regression, but not constrained to linear
- Classification
  - Find which class is the closest match
- Pattern Recognition
  - Fined tuned classification
- Self organizing map for clustering

◆ Not constrained to linear or Gauss Normal distribution

◆ Also used for modeling biological neural network in medical research

Advantages of Neural Network

- No Expert needed
- No Knowledge Engineer needed
- Does not have bias of expert
- Can interpolate for all cases
- Learns from facts
- Can resolve conflicts
- Variables can be correlated (multicollinearity)

More Advantages

- Learns relationships
- Can make good model with noisy or incomplete data
- Can handle non-linear or discontinuous data
- Can Handle data of unknown or undefined distribution
- Data Driven

Disadvantages of Neural Net

- Black Box
  - don’t know why or how
  - not sure of what it is looking at
- Operator dependent
- Don’t have knowledge in hand
- Just a mapping of input to output, no modeling of system

* *Many of these disadvantages are being overcome*

Black Box

- What happens inside the box is unknown
- We can’t see into the box
- We don’t know what it knows

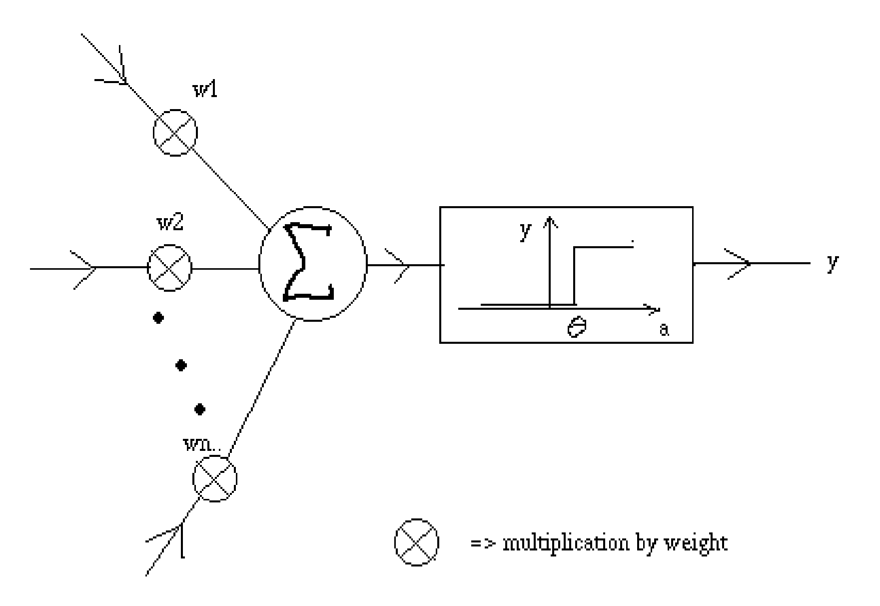
Biological Neural Network

- Human Brain has 4 x 10^{10} to 10^{11} Neurons
- Each can have 10,000 connections*
- Human baby makes 1 million connections per second until age 2
- Speed of synapse is 1 kHz, much slower than computer (3.0+ gHz)
- Massively parallel structure

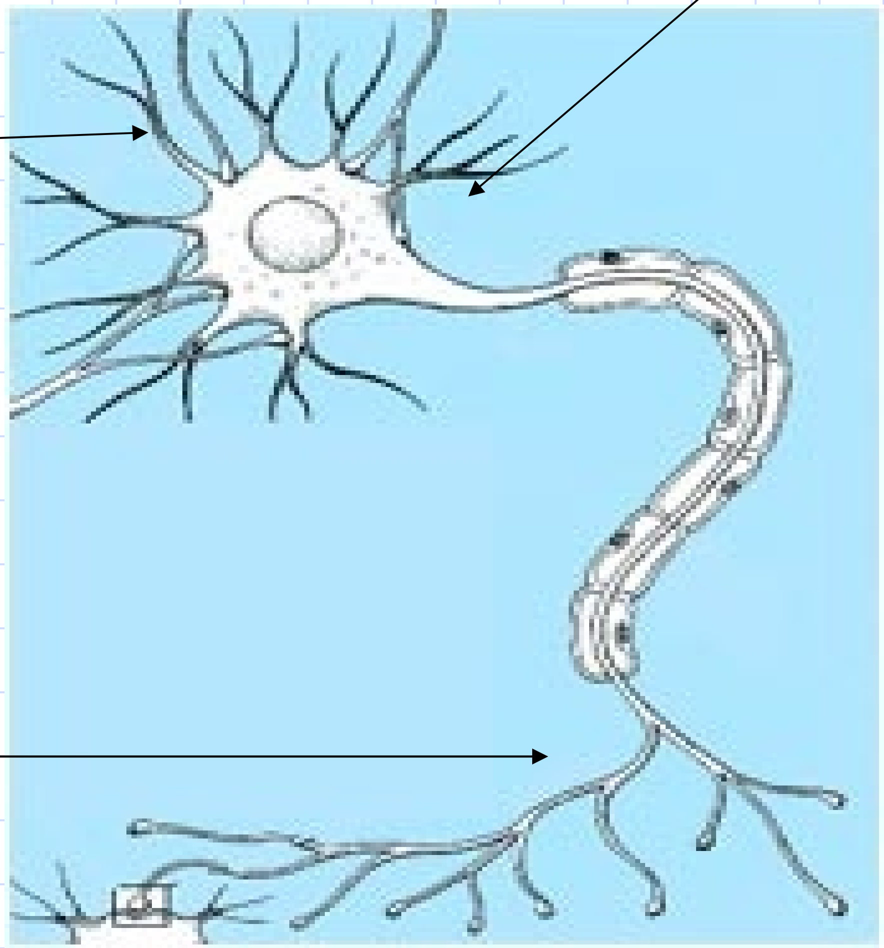
* Some estimates are much greater, as much as 100,000

How does a neuron work?

- It sums the weighted inputs
  - If it is enough, then neuron fires
  - There can be as many as 10,000 or more inputs



Neuron

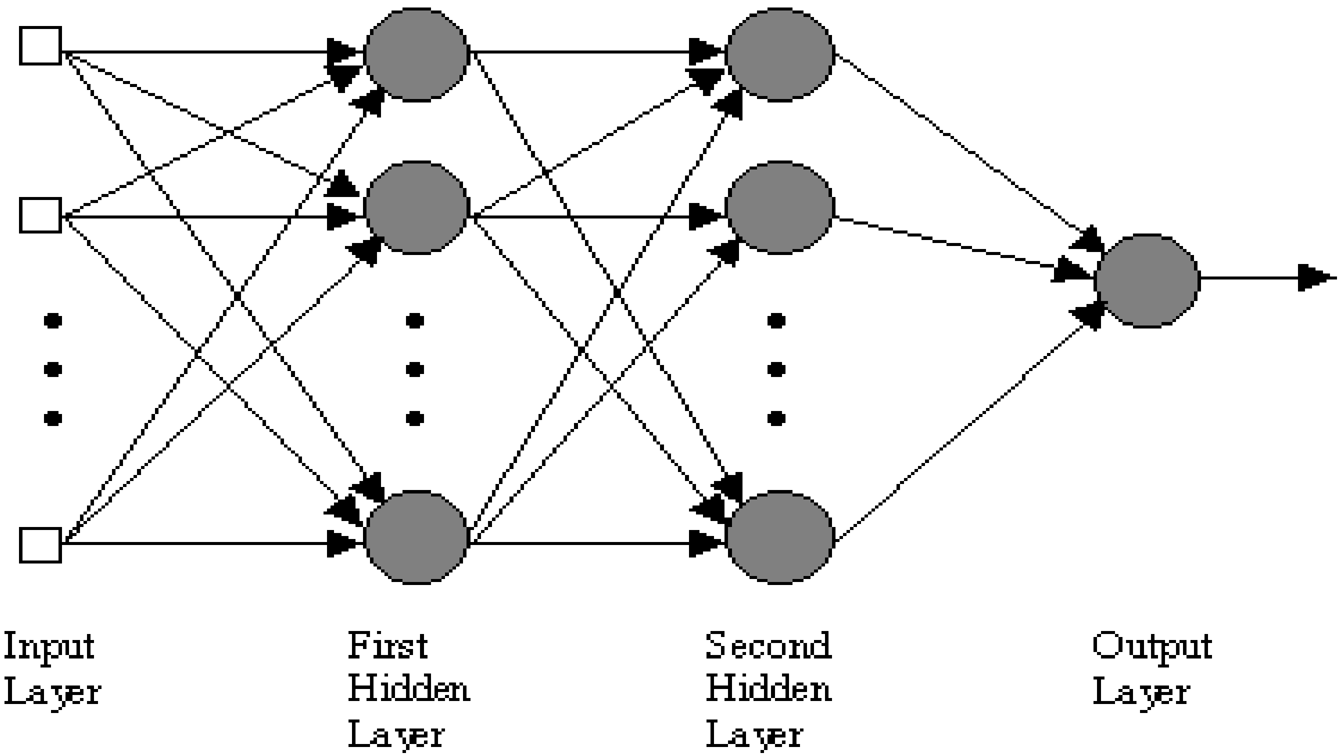


soma (body)

Dendrites
(inputs)

Axons
(outputs)

Neural Network



This is a feed-forward design

Computer Neural Network

- Von Neumann architecture
- Serial machine with inherently parallel process
- Series of mathematical equations
- Simulates relatively small brain
- Limited connectivity
- Closely approximates complex non-linear functions

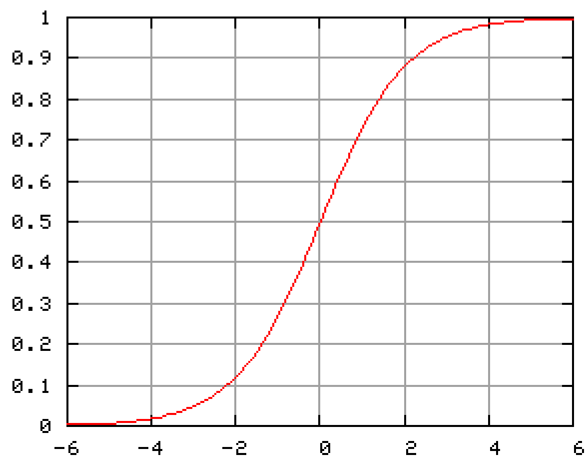
Neuron Activation

- Weights can be positive or negative
- Negative weight inhibits neuron firing
- Sum = $W_1N_1 + W_2N_2 + \ldots. + W_nN_n$
- If sum is negative, neuron does not fire
- If sum is positive neuron fires
- Fire means an output from neuron
- Non-linear function
- Some models include a threshold

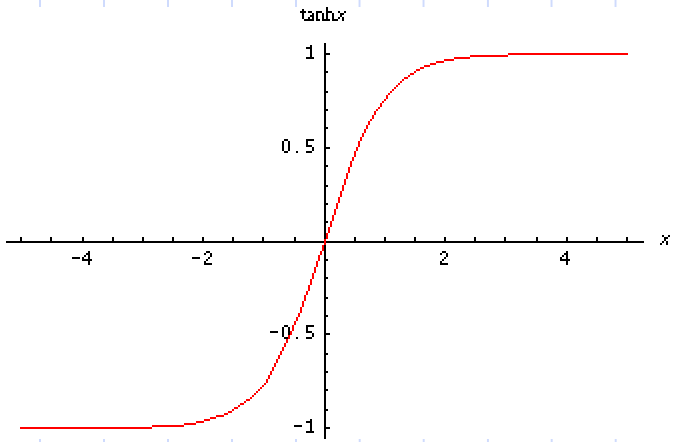
Neuron Activation

- Linear
- Sigmoidal
  - $1.0/(1.0+e^{-s})$ where $s = \Sigma$ inputs
  - 0 or +1 result
- Hyperbolic Tangent
  - $(e^{s} - e^{-s}) / (e^{s} + e^{-s})$ where $s = \Sigma$ inputs
  - -1 or +1 result
- Also called squashing or clamping function
  - Because it takes a large value and compresses it
  - Adds the non-linearity to the process

Activation Functions



Sigmoidal Function
Goes from 0 to 1
Hard to be at extreme



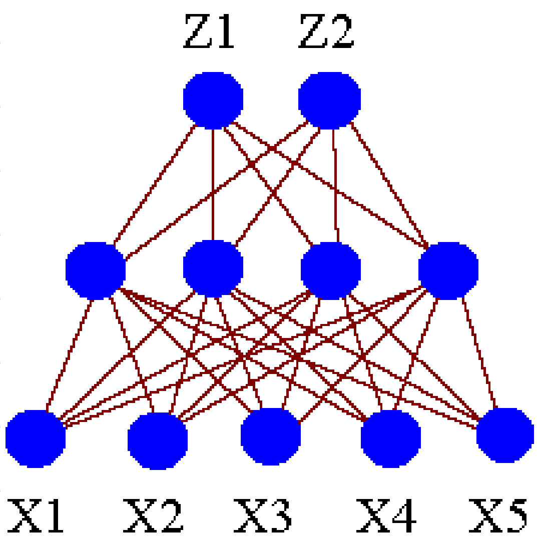
Hyperbolic Tangent
Goes from -1 to 1
Hard to be at extreme

Neuron Math

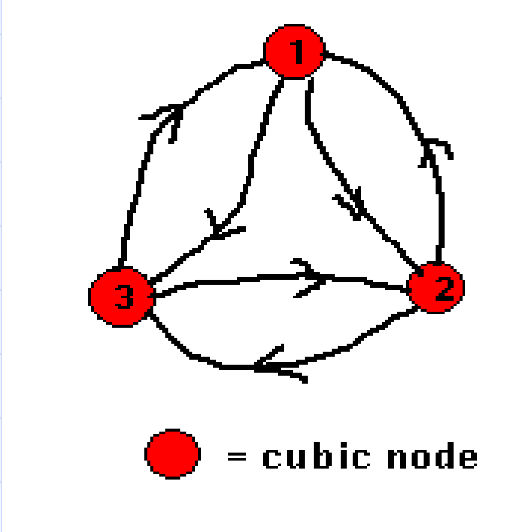
- Don't try for 0 or 1
  - Use 0.1 and 0.9 instead for logistic
  - Use –0.9 and +0.9 for hyperbolic tangent
- Real plane math
- Complex domain math
  - Quite often outperforms systems using real domain math
  - Better for signal & image processing

What does the network look like?

- This is a computer model, not biological
- Left has 11 neurons, sea slug has 100

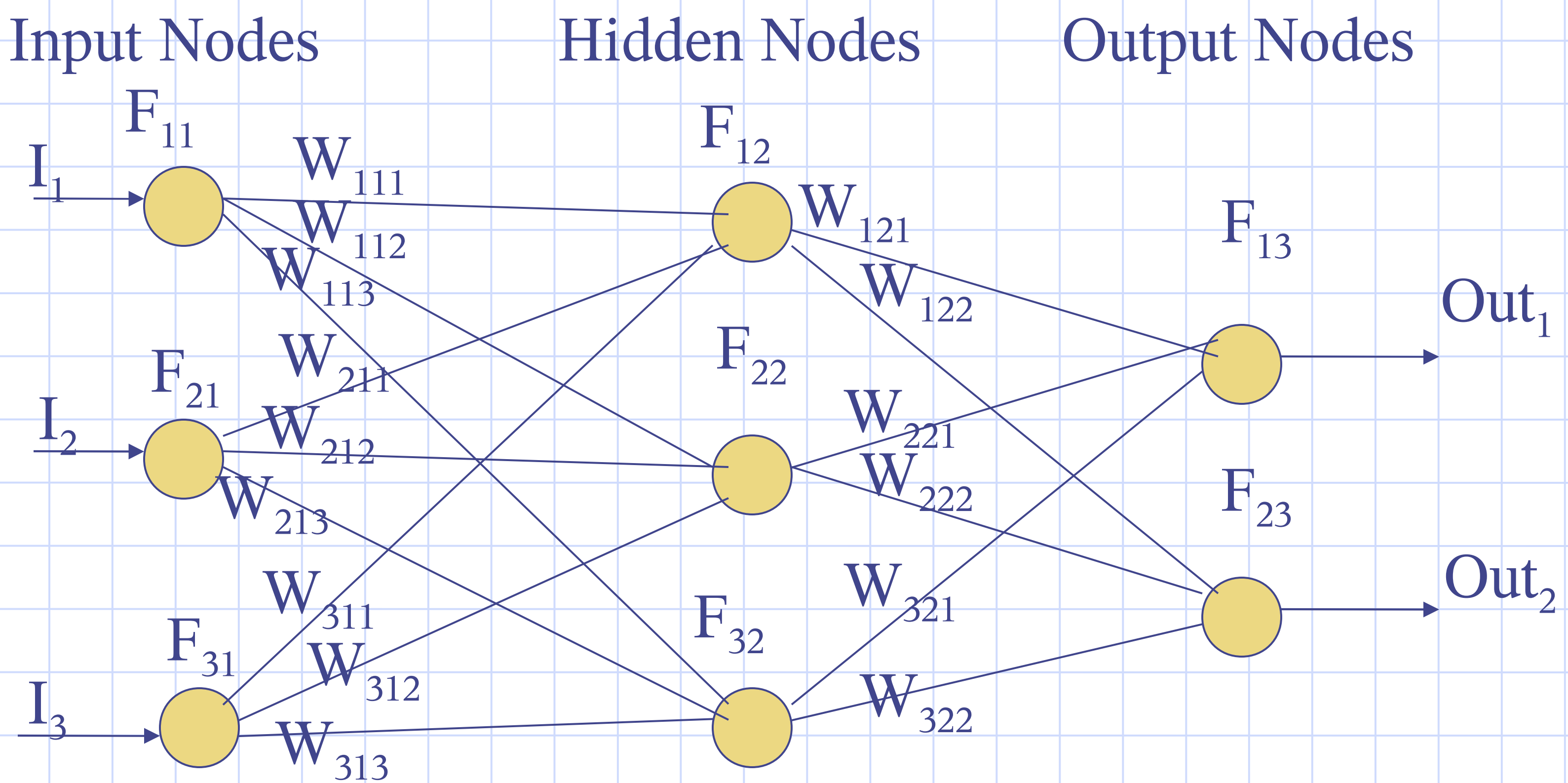


Feed Forward



Recurrent or Feedback

Small Neural Network



Regression?

- With linear activation, this is but parallel regression
- With sigmoid or H-Tan, this is a parallel logistic regression
- An ANN with zero hidden nodes, one output, and linear activation ***is*** OLS regression if the objective function is minimizing SSE (sum of squared error)

Mathematical Equations

- Input to $Hidden_{12} = H_1$
- $H_1 = [(I_1 * F_{11}) * W_{111}] + [(I_2 * F_{21}) * W_{211}] + [(I_3 * F_{31}) * W_{311}]$
- $H_2 = \ldots\ldots$
- $H_3 = \ldots\ldots$
- $Out_1 = [(H_1 * F_{12}) * W_{121}] + [(H_2 * F_{22}) * W_{221}] + [(H_3 * F_{32}) * W_{321}]$

Matrix Math

- Makes it very simple!

F(A x W) = Out

In Fortran:

out = Active(matmul(input, weights))

Where F or Active is the activation function

Can also use Matlab/Mathematica but it will compute more slowly as they are interpretive

Comparison to Regression

- OLS with 3 independent and 1 dependent variables would have a maximum of 3 coefficients and 1 intercept
- With 2 dependent variables, it would require Canonical Correlation (general linear model) and the same number of coefficients
- ANN (with one hidden layer) has 15 coefficients (weights) and activation functions can be non-linear
- Multicollinearity is not a problem in an ANN

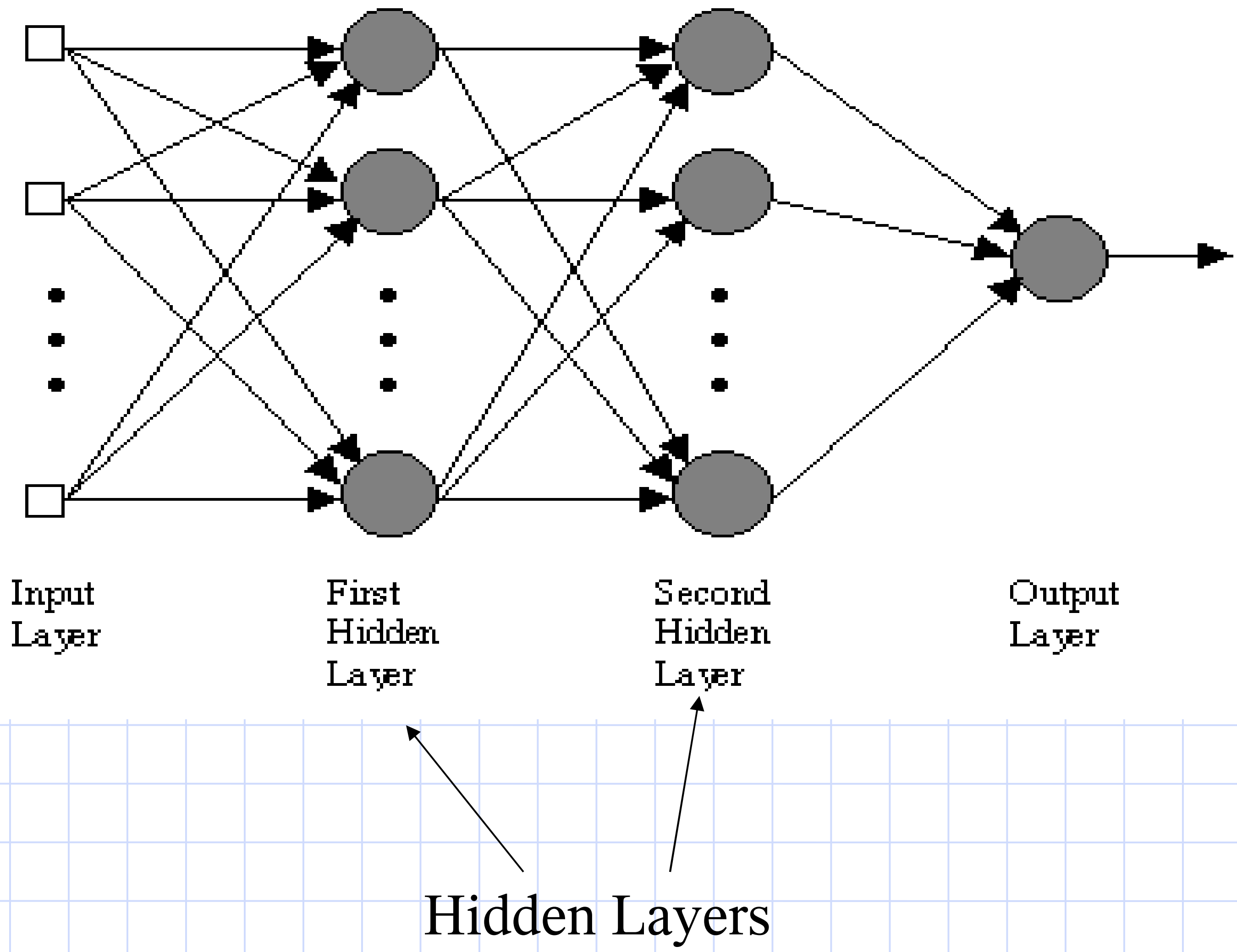
Inputs

- One per input node
- Ratio
- Logical
- Dummy
- Categorical
- Ordinal
- Fuzzy (PNL)
- Functional Link Network
  - Interaction variable
  - Transformed variable

Hidden Layer(s)

- Increase complexity
- Can increase accuracy
- Can reduce degrees of freedom
  - Need larger data set
- Presently architecture up to programmer
- Source for error
- In future will be more automatic
  - Some literature describes this

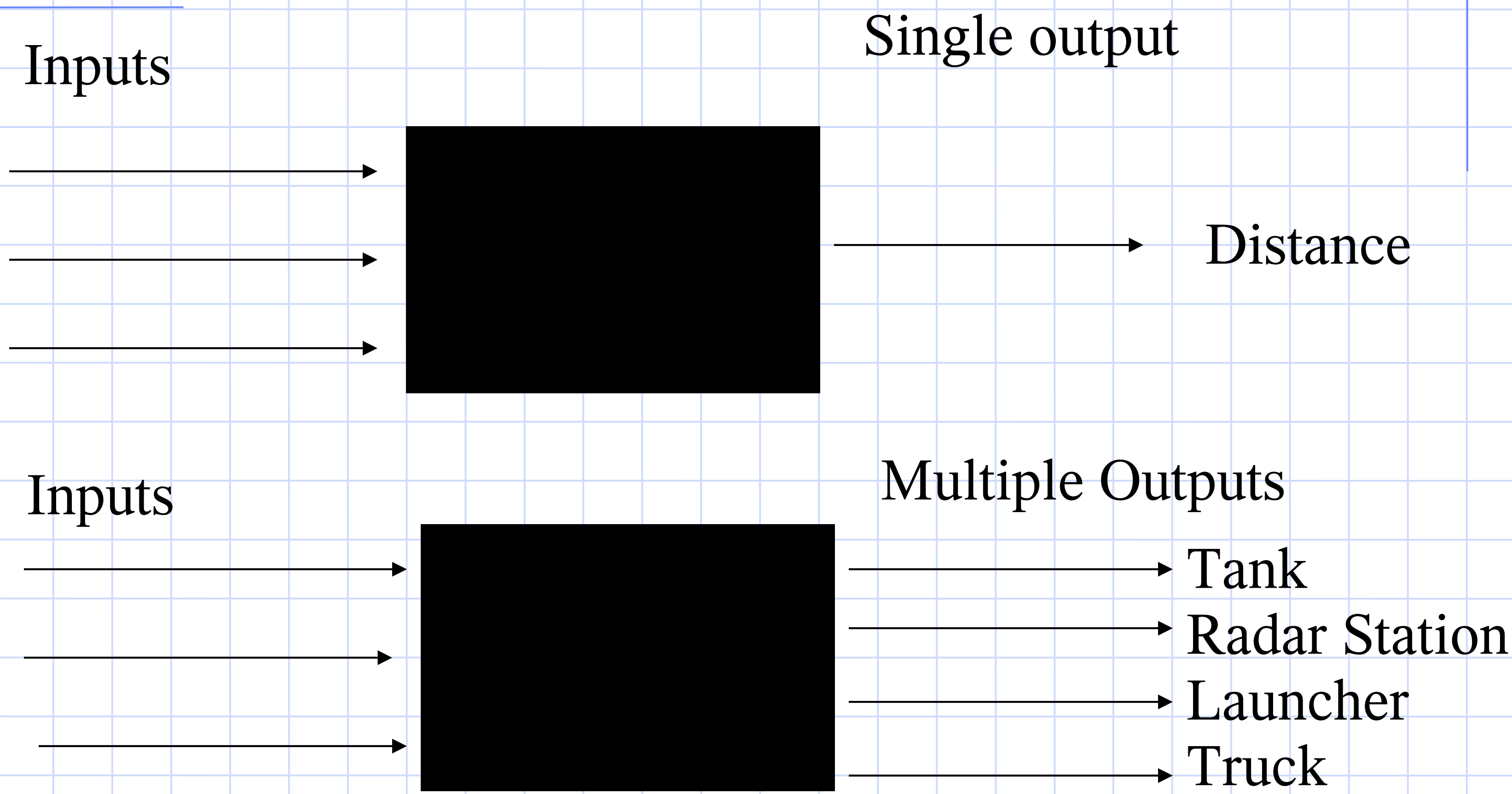
Hidden Layer(s)



Outputs

- One for single dependent variable
- Multiple
  - Prediction
  - Classification
  - Pattern recognition

Outputs


Inputs
Single output
Distance
Inputs
Multiple Outputs
Tank
Radar Station
Launcher
Truck

Macro View of Training

- Setting all of the weights
- To create optimal performance
- Optimal adherence to training data
- Really an optimization problem
  - Optimal methods depends on many variables
  - See optimization lecture
- Need objective function
- Beware of local minima!

Supervised or Not

- Supervised
  - Train it with examples
  - And give it the answers
  - Much like school
- Unsupervised
  - Give it examples
  - Do NOT give it answers
  - It organizes the data by similarities

Training

- Supervised

Pattern 1 → Answer 1
Pattern 2 → Answer 2
Pattern 3 → Answer 3

- Unsupervised

Pattern 1
Pattern 2
Pattern 3

Optimization Methods to Set the Weights

- Back Propagation (most popular)
- Gradient Descent
- Generalized reduced gradient (GRG)
- Simulated Annealing
- Genetic Algorithm
- Two or more output nodes
  - Multi objective optimization (hard problem)
- Many more methods

Training Data Set

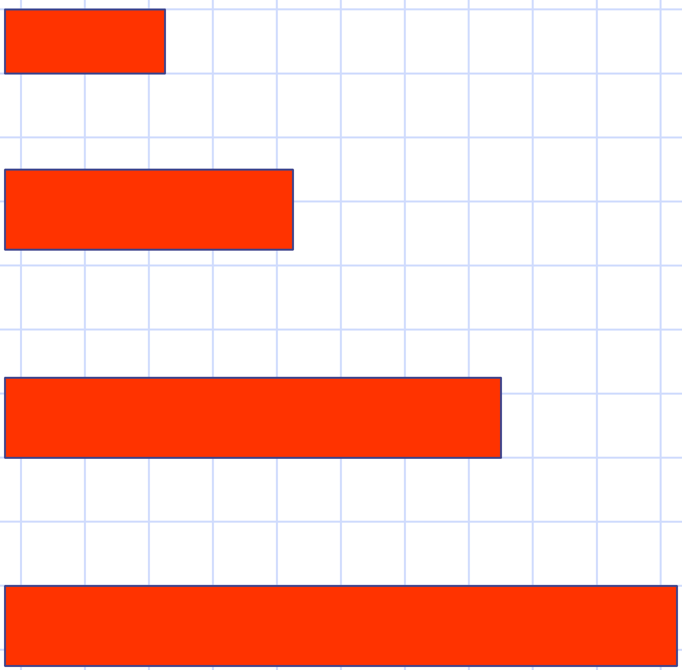
- Need more observations than weights
  - Positive number degrees freedom
- More observations is usually better
  - Lower variance
  - More knowledge
- Watch aging of data
- Data must be representative of population

Data Window

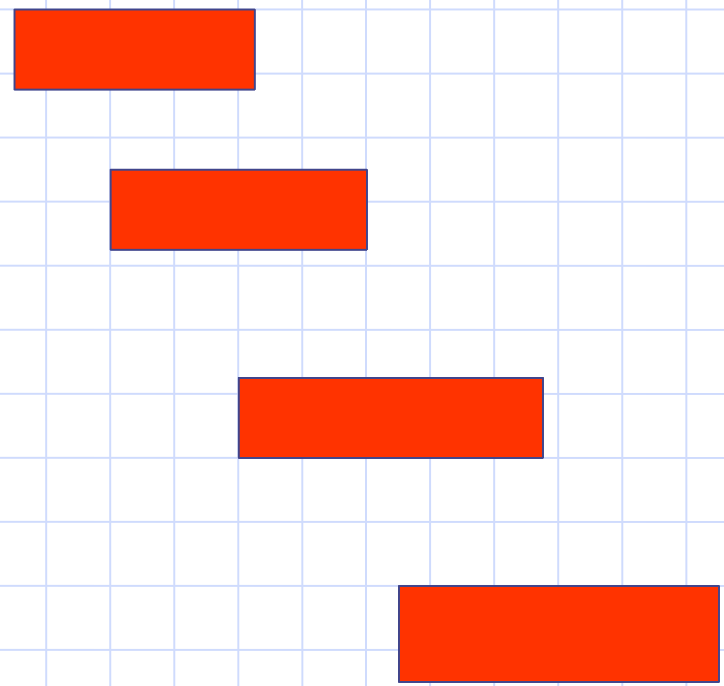
- Rolling Window
  - Rolls forward including all data behind
  - Constant starting point with ever increasing size
- Moving Window
  - Deletes the oldest as it adds the newest
  - Constant size with ever increasing starting point
  - Necessary when underlying factors change

Rolling vs. Moving Window

- Rolling Window



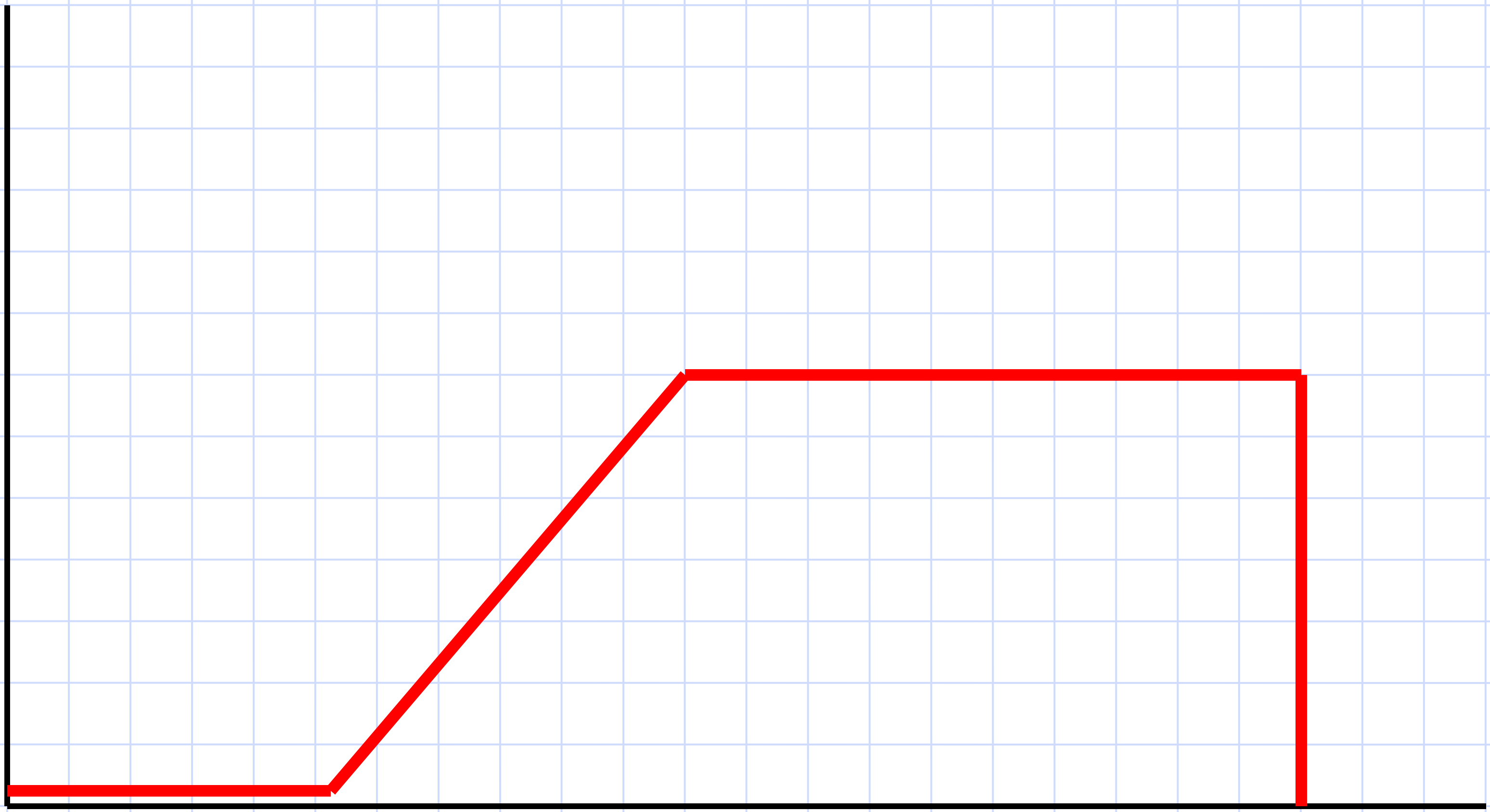
- Moving Window



Data Window Continued

- Weighted Window
  - Morantz, Whalen, & Zhang
  - Superset of rolling & moving window
  - Oldest data is reduced in importance
  - Has reduced residual by as much as 50%
  - Multi factor ANOVA shows results significant in majority of applications with real world data

Weighted Window



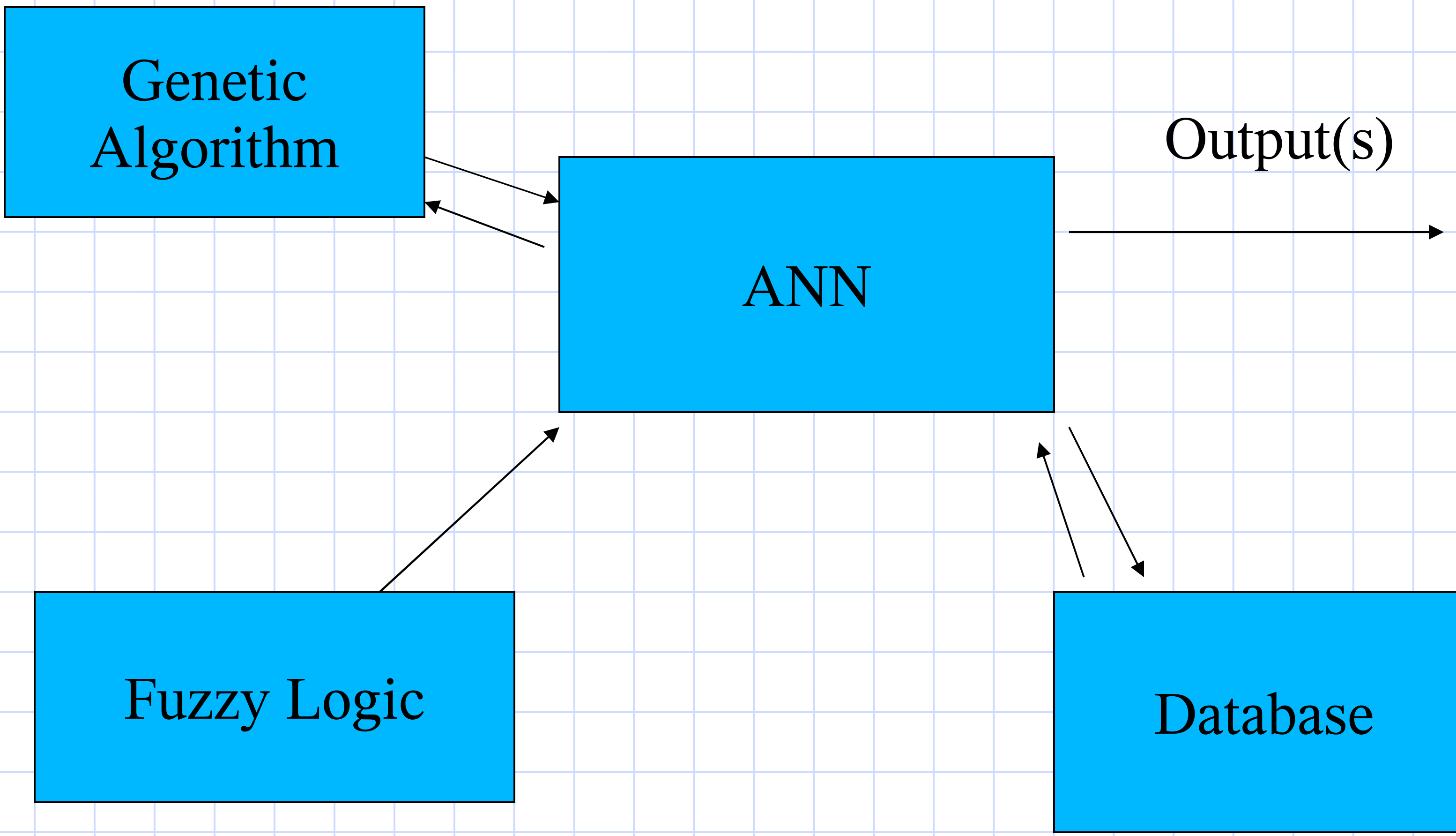
Dynamic Learning

- Also called reinforcement learning
- Continuous learning
  - From mistakes and successes
  - From new information
- Shooting baskets example
  - Too low. Learned: throw harder
  - Too high. Learned: throw softer, but not as soft as before
  - Basket! Learned: correct amount of “push”
- Loaning $10 example

Hybrids

- Combine several systems
  - GA and ANN
  - ANN with fuzzy, GA, & database
  - Many possibilities
- Uses more methods than just one type
- Can seed system with expert knowledge and then update with data
- Sometimes hard to get all parts to work together
- Harder to validate model

Hybrids



Example

- You go some place that you have never been before, and get “bad vibes”
  - Atmosphere, temperature, lighting, smell, coloring, numerous things
- For some reason, brain associates these together, possibly some past experience
- Gives you “bad feeling”

Additional Examples

- Military: submarine, tank, & sniper detection
- Security
- Classify stars & planets
- Data mining
- Natural language recognition
- OCR including Kanji

My Favorite Examples

- Fire control for ABL (air borne Laser)
- ANN with GA hybrid
- With real constraints
- Initially trained from panel of experts
- Ran in simulation
  - Learned from mistakes
  - Retrained after each set of sorties
  - Improved performance (less leakers)
  - From Stroud, IEEE Transactions on Neural Networks

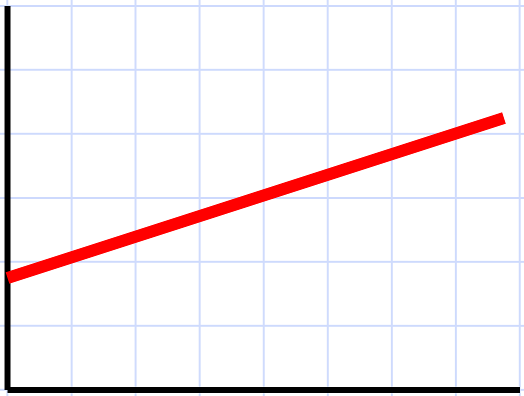
The Other Favorite Example

- The brain of a bat
  - Size of a plum
  - Controls voluntary & involuntary processes
  - Controls sonar system and navigation
    - Outperforms our best navigation systems
    - Bat can fly through moving electric fan

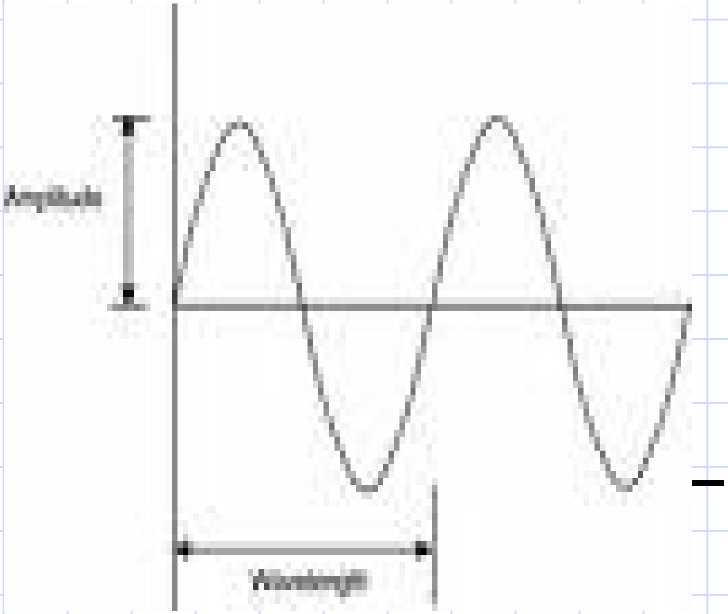
When to Use?

- Look at the data
- Is data linear over range of interest?
- Is Regression accurate enough?
  - Occam's razor says to use it if it is
- Is data non-linear and/or discontinuous?

What to Use



Regression is fine



Use the ANN here
Regression won't
Fit it well

ANN Chip

- Original funding was from TEAMA
  - Goal was for use as intelligent appliance
    - Toaster learned how you like your toast
    - Coffee pot learned how you want coffee
- JPL
  - Stack chip
  - For vision applications

Future

- Rule extraction
- Hybrids
- Dynamic learning
- Parallel processing (it is here)
- Dedicated chips (ZISC chip)
- Bigger & more automatic
- Machine Cognition

About Me

- I am a Decision Scientist
- I work on methods to make intelligent ***High Quality*** decisions
- Neural networks are a tool in my tool-box
- I use them like regression, except that they can be non-linear
- Not the case of only having a hammer and all problems looking like a nail.

Information Sources

- www.machine-cognition.com
- IEEE Transactions on Neural Networks
- IEEE Intelligent Systems Journal
- IEEE Computational Intelligence Society
- AAAI American Association for Artificial Intelligence
- www.ieee.org
- Internet

Contact Info

- www.machine-cognition.com
- bradscientist@ieee.org