Neural Networks
More than you ever wanted to know

Presented to the
Phoenix IEEE Computer Society
10 February 2016

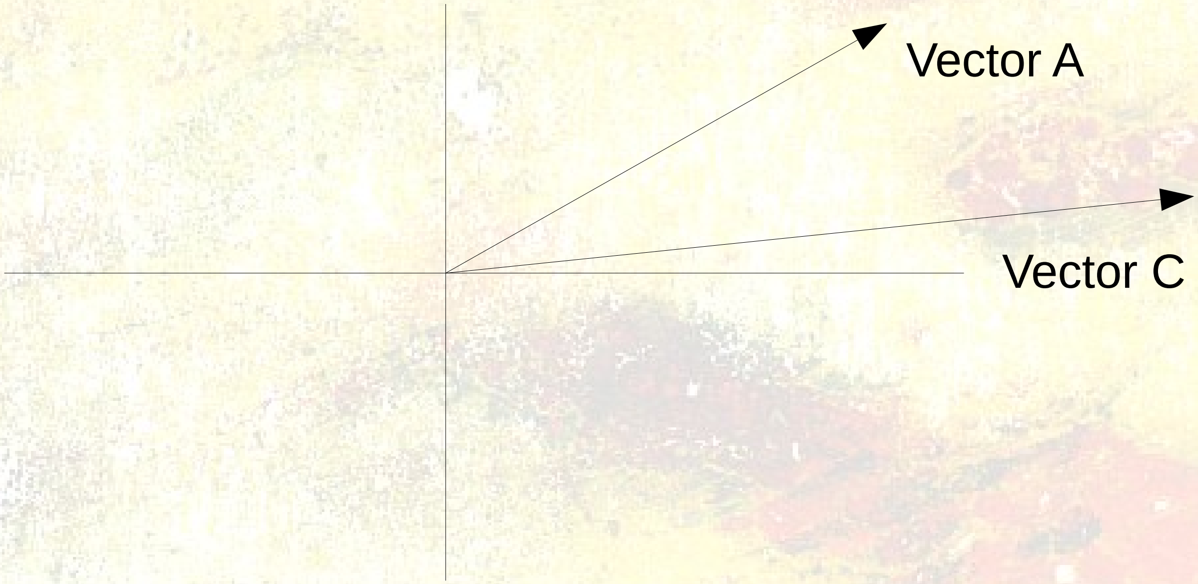
Contents

Overview & Focus

- I am a decision scientist
- Focus is on knowledge & decision-making
- How various life forms make decisions
- Biologically inspired computing (BICA)
- In addition to mathematics
- Biological brains are neural networks
- So is central nervous system (CNS)
- Artificial Neural Networks (ANN)

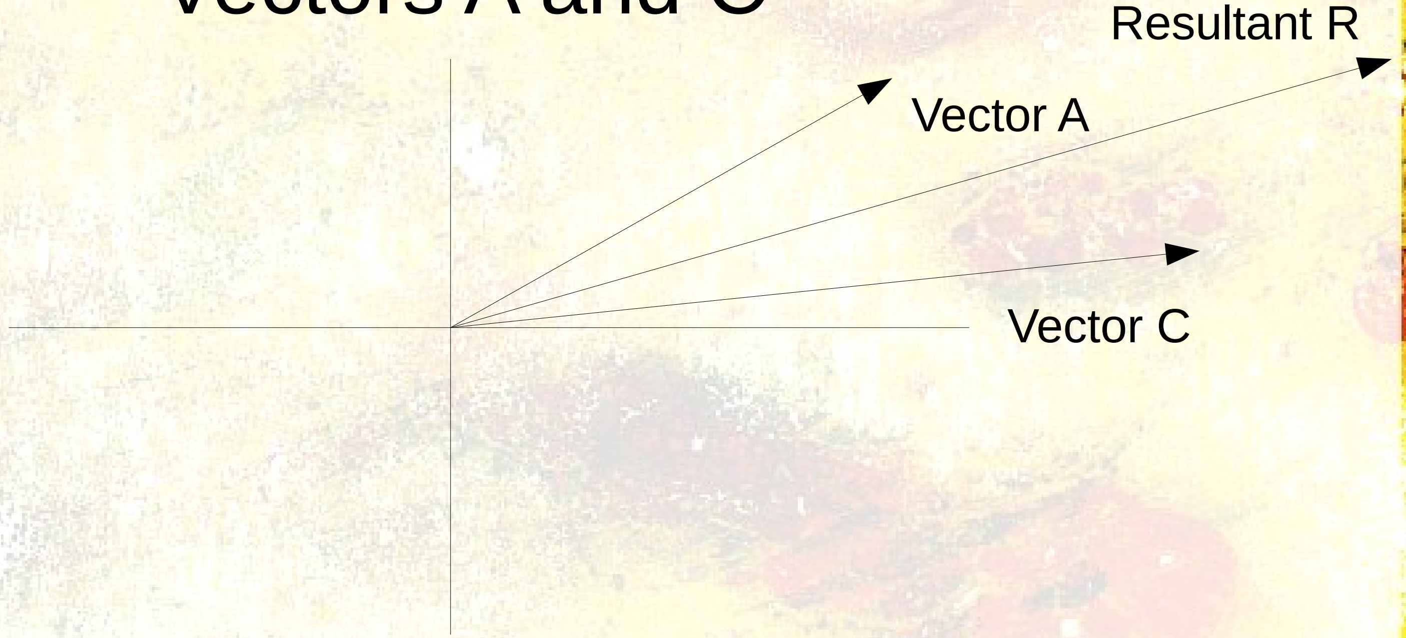
Vector Math

Vectors A and C



Vector Math

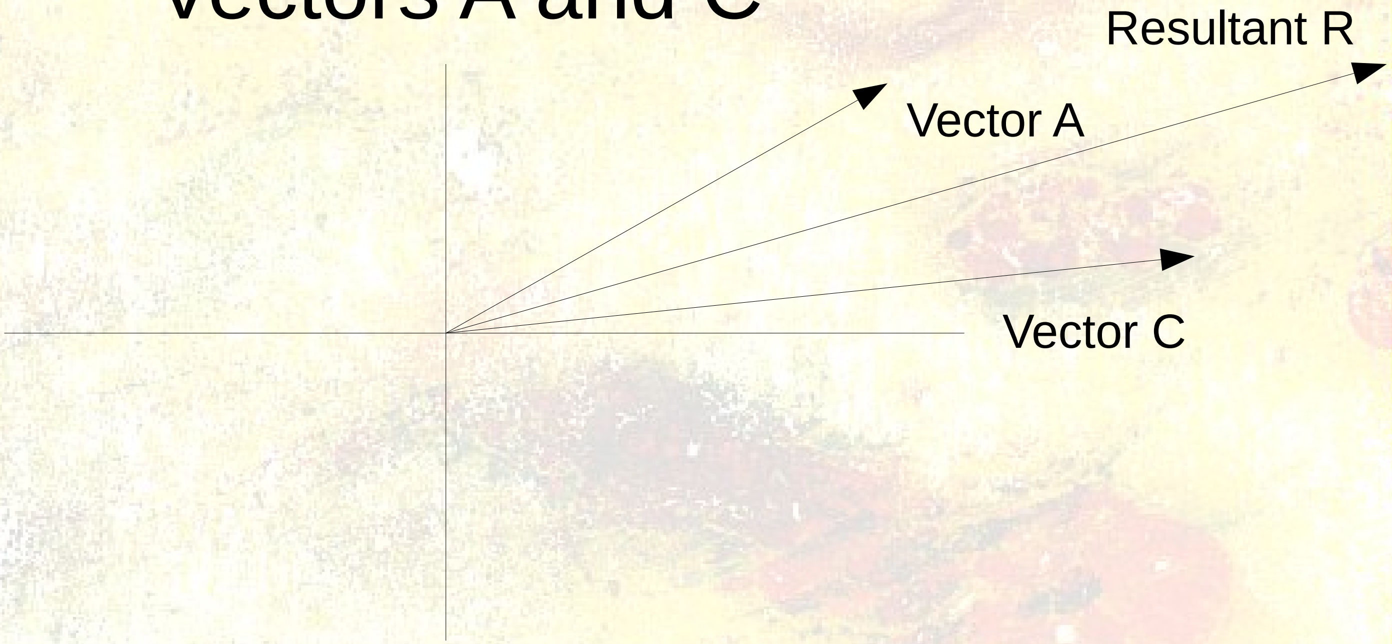
Vectors A and C



This is very important in Electronics

Vector Math

Vectors A and C



This is an RCA Vector

Vector Math Explained

RCA VICTOR

Degrees of Freedom

- The number of ways by which a dynamic system can change without violating any constraint imposed on it
- The number of values that are free to vary
- e.g. 3X + 2Y = 14 (1 equation, 1 variable)
  - There is one df (degree of freedom)
  - If also had 2X + 4Y = 6 (2 equations, 2 variables)
  - Then no degrees of freedom
    - X and Y are fixed
  - If also had 5X + 4Y = 17 (3 equations, 2 variables)
    - Negative number df
    - May not even be solvable

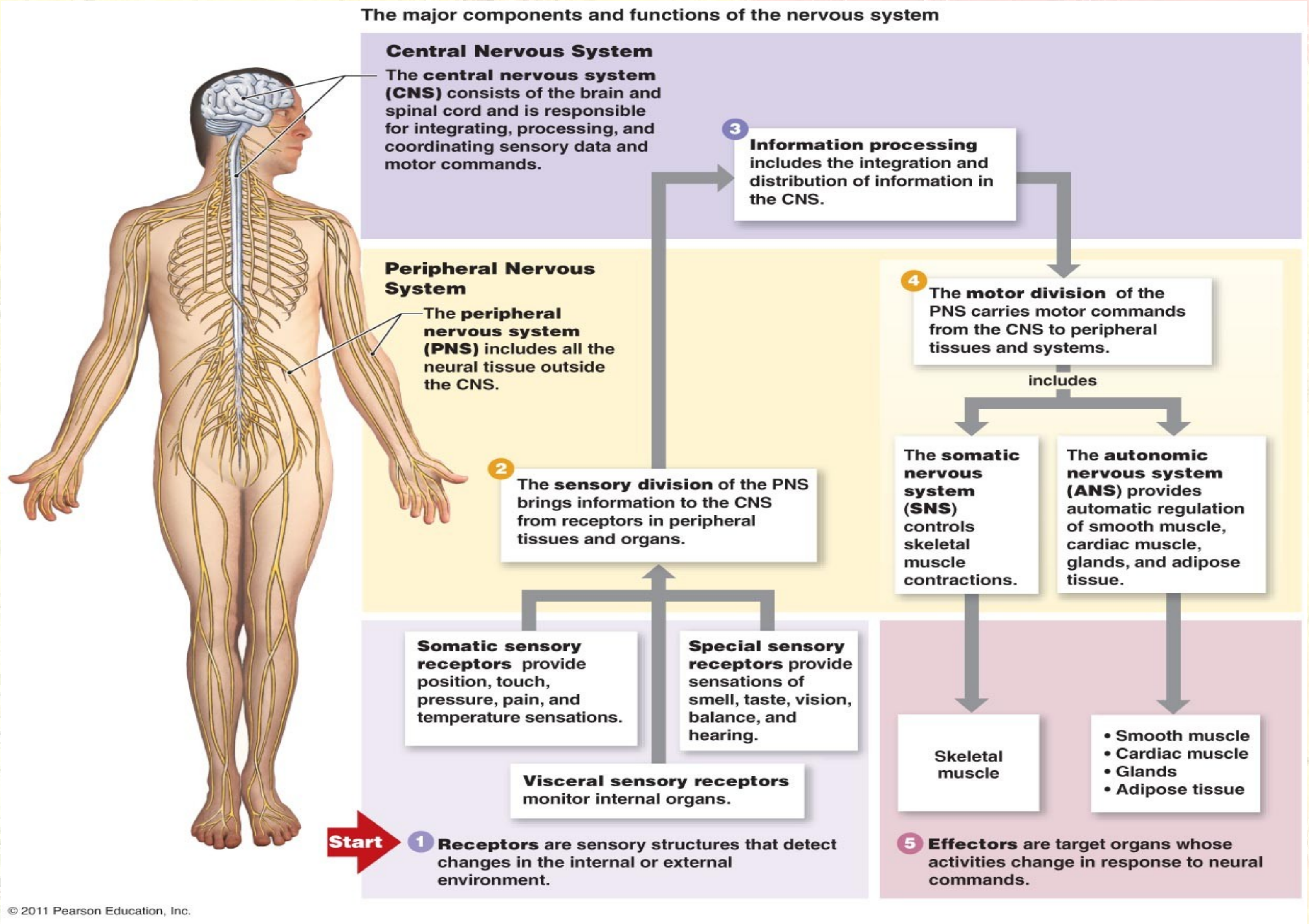
Amazing Animals

- Dolphins
- Koala bears
- Lions
- Dogs
  - Bloodhounds
- Eagles
  - Vision
- Homo sapiens
- Monkeys
- Lemurs
- Bats
  - Echolocation
- Homing pigeons
  - Location finding

What is the Organ of Vision?

- The brain!
- The ears, eyes, nose are all sensors
- Image/decision is made in the brain
- All connected by the central nervous system (CNS)
- There is also preprocessing in the retina

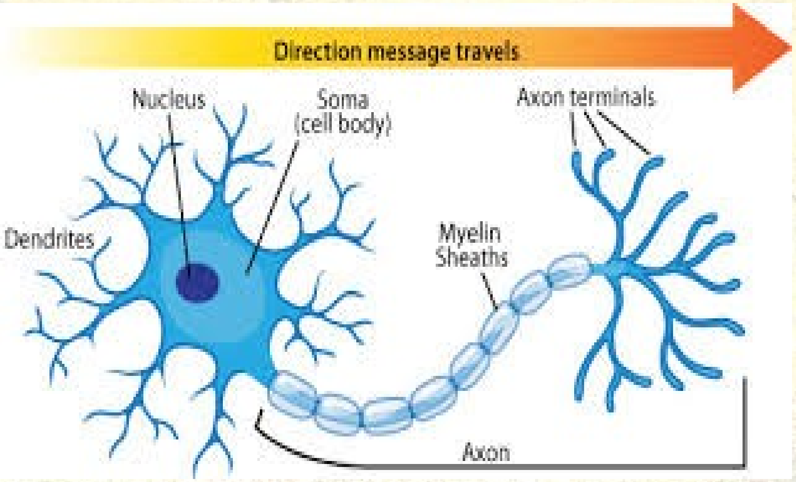
Central Nervous System

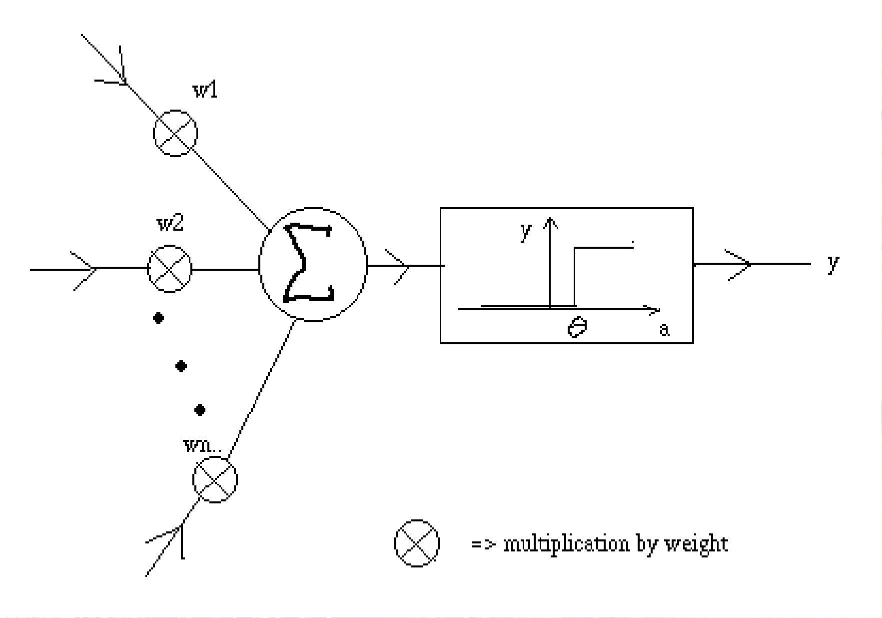


Beginning Research

- Hodgkin & Huxley 1952
  - Giant sea snail (aplysia calfornica)
  - Giant squid axon
  - Laws about movement of ions in nerve cells during action potential
- Before electron microscope
  - Formulas for nerve cell membrane
  - Before they could be seen

Neuron





Drawing

Mathematical view

It sums the weighted inputs

1940's

- McCullough & Pitts
  - Computational model
    - Threshold logic
    - 1943
- Donald Hebb
  - Psychologist
  - Hebbian learning
  - Unsupervised
  - Self learning

More History

- Rosenblatt
  - Perceptron
  - 1958
- Minsky & Papert
  - Can not do XOR
  - Too computationally intensive
- Werbos
  - Back Propagation
  - 1975
- Electromechanical version

Renaissance

- Rumelhart & McClelland
  - Parallel Distributed Processing
  - 1986
- PCs and more power in computers
- Other ways to set weights
- Lots of speed and computing power
- Artificial Neural Networks (ANN) come back

Summary of Biological NN

- Electro-chemical process
- Has effective clock speed of 1 Khz
- Stores knowledge in the connection weights
- We do NOT know how it sets the weights
- We do NOT understand creativity or intuition
- We do NOT understand ESP, psycho-kinesis
- It does have massive parallel operation
- We do NOT understand timing along axons

Interesting Tidbits

- A baby, until age 2, develops 1 million interconnects per second
- Average human has between 4E10 to 1E11 neurons
- Each neuron can have 1E4 interconnects
  - Some claim 1E5 (do the math)

What is an ANN

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

What Can an ANN Use to Make Connections/Mapping?

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

Four types of Functions

- Prediction and Time Series Forecasting
  - Like regression, but not constrained to linear
- Classification
  - Sort into a class, like cluster analysis
- 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)
- Does not need linear or Gauss Normal

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

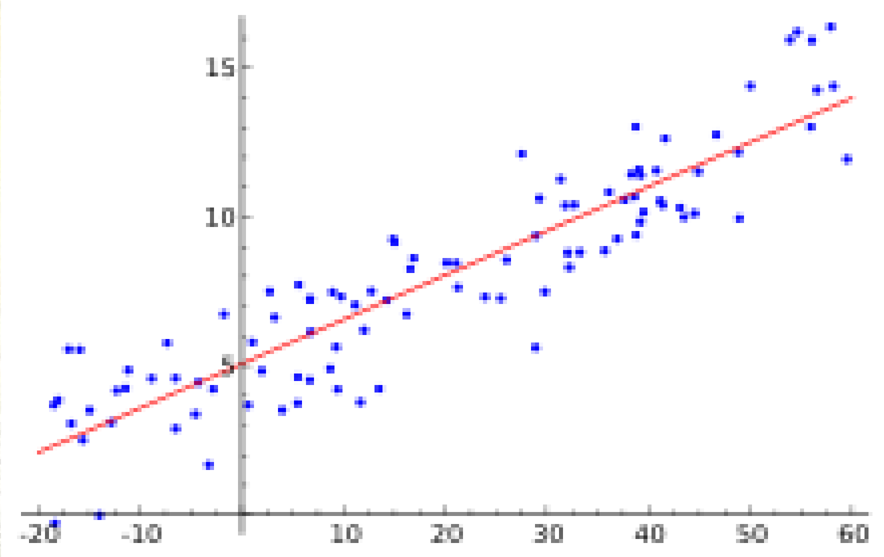
- Black Box
  - don’t know why or how
  - not sure of what it is looking at
- Operator dependent
- Don’t have knowledge in hand
- ** Many of these disadvantages are being overcome*

Black Box

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

Regression

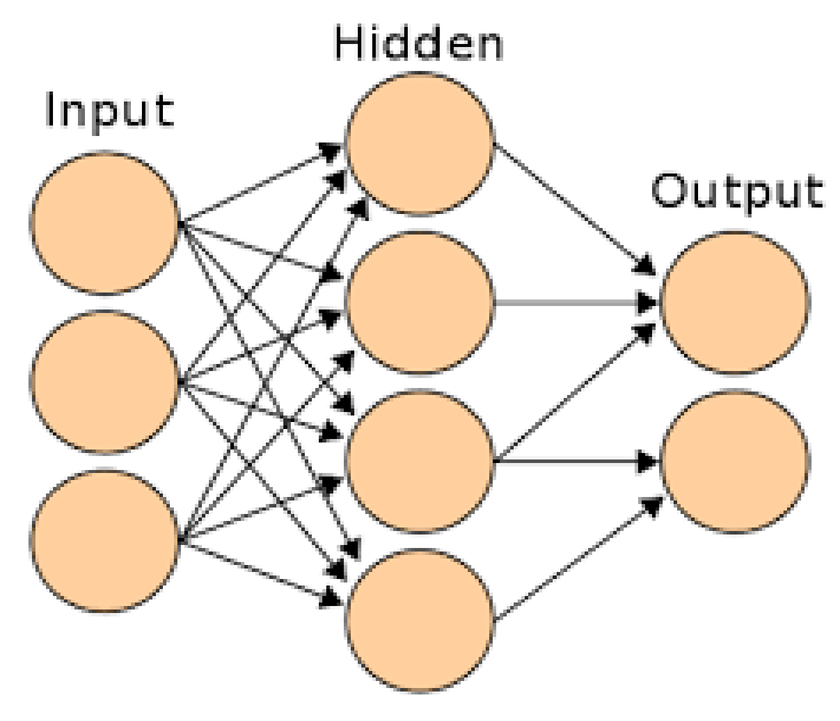
Linear Regression



Y = aX + bY +c is equation of the line

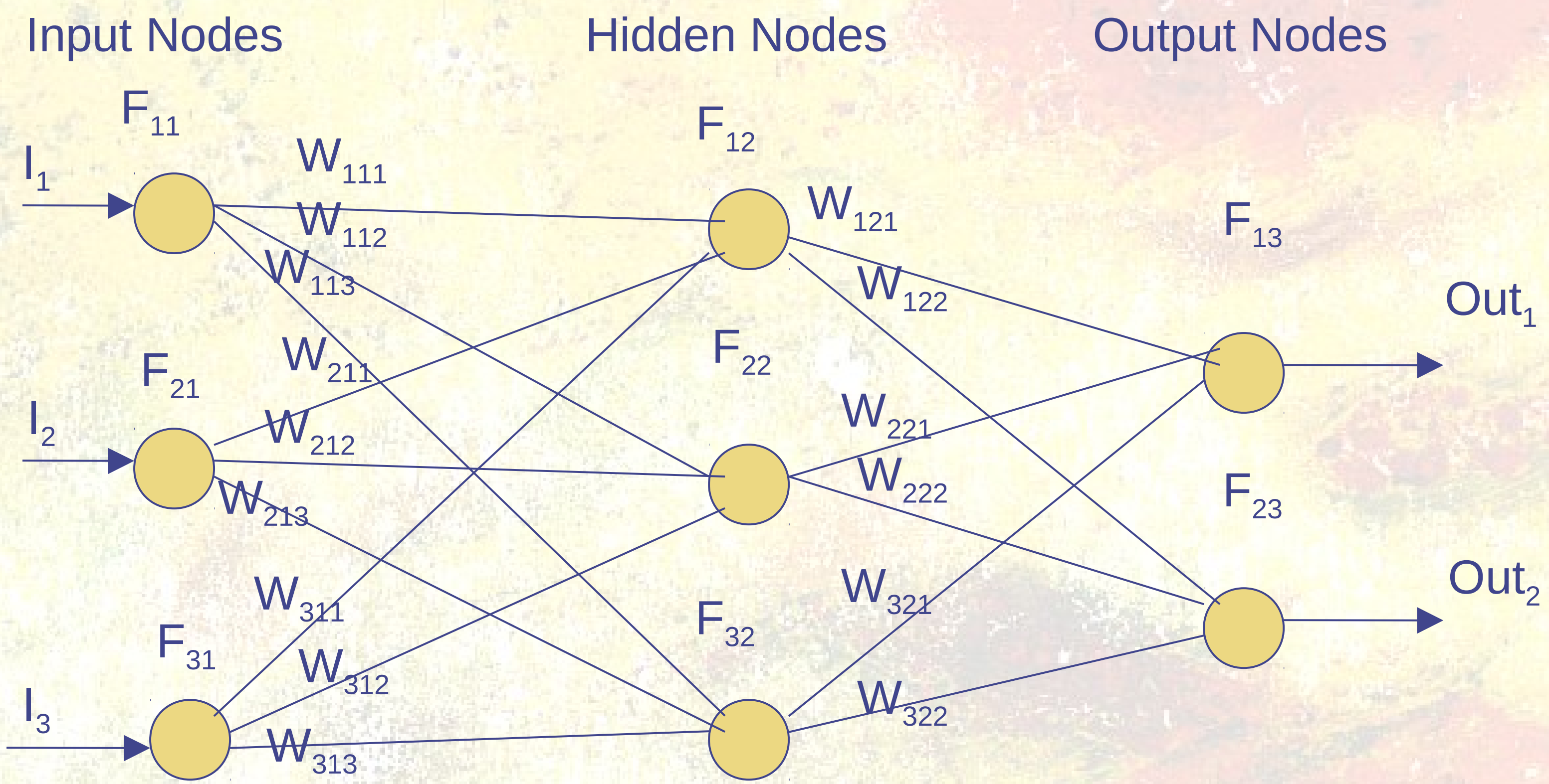
The dots are the real data points

Suppose



You had a whole lot more equations & coefficients
You were not limited to linear math (activation functions)
But there is a loss of degrees of freedom
Can have multiple hidden layers ("Deep learning")

Small Neural Network



See how many equations there are describing the system

Mathematical Equations

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

With a few more neurons, it becomes many more equations
Think about another hidden layer

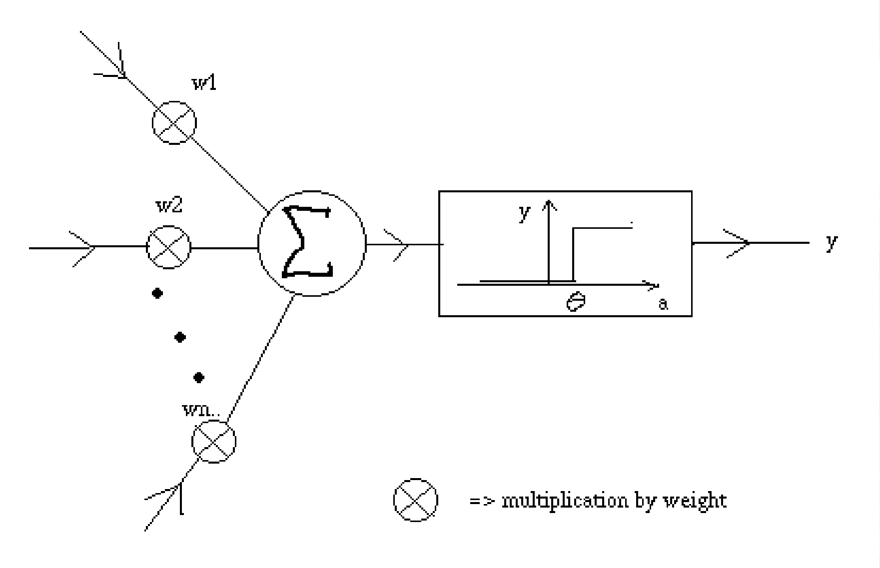
Linear Algebra to the Rescue

- For each layer
  - A matrix of the weights
  - A matrix of the inputs
- An Activation function called 'Active'
- output = Active(matmul(input, weights))*
  - Where Active is the activation function

* This is in Fortran 90 (and up)
Can also do this in Octave or Matlab

What The Neuron Does

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



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 (over threshold) neuron fires
- Fire means an output from neuron
- Non-linear function
- Some models include a threshold

Activation Functions

- 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

Neuron Math

- For other than Linear
- 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
  - Squares up the corners
- Real plane math
- Complex domain math
  - Quite often outperforms systems using real domain math
  - Better for signal & image processing
- Need to scale values

Sum Total

- Huge equation
- Tons of coefficients
- Non-linear activation functions
- Allows it to fit the data
- Never understands what is going on
- Just fits the data
- Contains knowledge

Training

- Like going to school & learning
- Setting the connection 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 Unsupervised

- 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
  - Self discovery

Training Methods

- Back Propagation (most popular)
- Gradient Descent
- Generalized reduced gradient (GRG)
- Simulated Annealing
- Genetic Algorithm
- Two or more output nodes
    - Multi objective optimization
- Many more methods
- Bio NN does it more efficiently, but we do not know how it does it

Training Data Set

- Need more observations than weights
  - Positive number degrees freedom
  - If not, use boosting* or bagging**
- More observations is usually better
  - Lower variance
  - More knowledge (the real key)
- Watch aging of data
- Data must be representative of population

* Singer, Schapire, & Freund

** Breiman

My Contribution

- Recency weighted ANN
- Time is a variable
  - Life & things change over time
  - Things in near future are more like what happened in near past
- It trains on all data but near past is more important, and at some point stops using old data
- This reduced residual >50% on some data sets

Recency Weighting Continued

- Did a two factor (blocked) ANOVA
- Compared Regression to ANN
- Recency weighting helped the ANN, not regression
- The ANN stores knowledge
- Regression builds a model

Dynamic 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

Biological 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”
- Intuition?

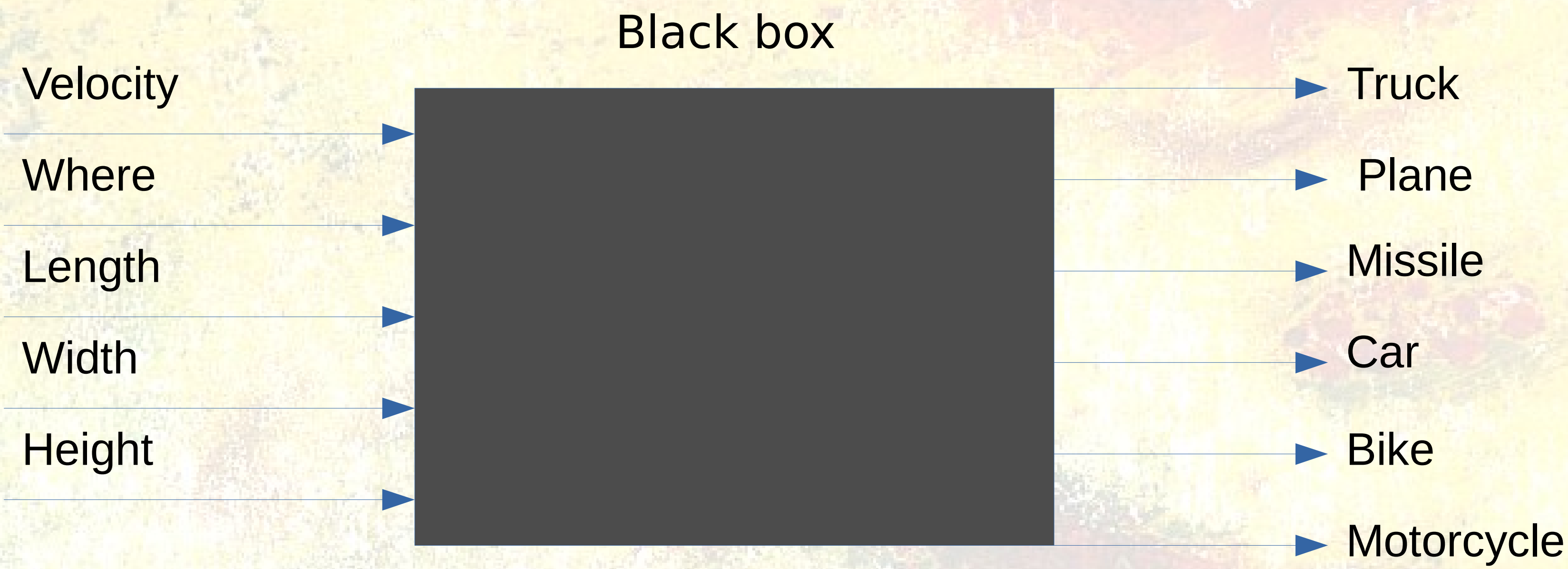
Computer Examples

- Military: submarine, tank, & sniper detection
- Security
- Classify stars & planets
- Data mining
- Natural language recognition
- OCR including Kanji
- A classifier in an ensemble learner

ABL Fire Control Example

- 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

Vehicle Classification Example



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
- Could use this same methodology for designing a missile, a flying saucer, etc

ANN vs Regression

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

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
- Many good books
- Internet

Thank You

Any Questions?