What is Deep Learning

This section describes the fields of Artificial Intelligence, Machine Learning and Deep Learning and how they are related.

What is Deep Learning?

The relationship between Deep Learning and Machine Learning

- Machine Learning
- Data Science / Data Mining
- Deep Learning

What is Machine Learning?

- A method of solving problems that can not easilly be solved through hand written code. Instead build a program that "learns" or adjusts itself to attain the correct answer.
- A program that creates a program.

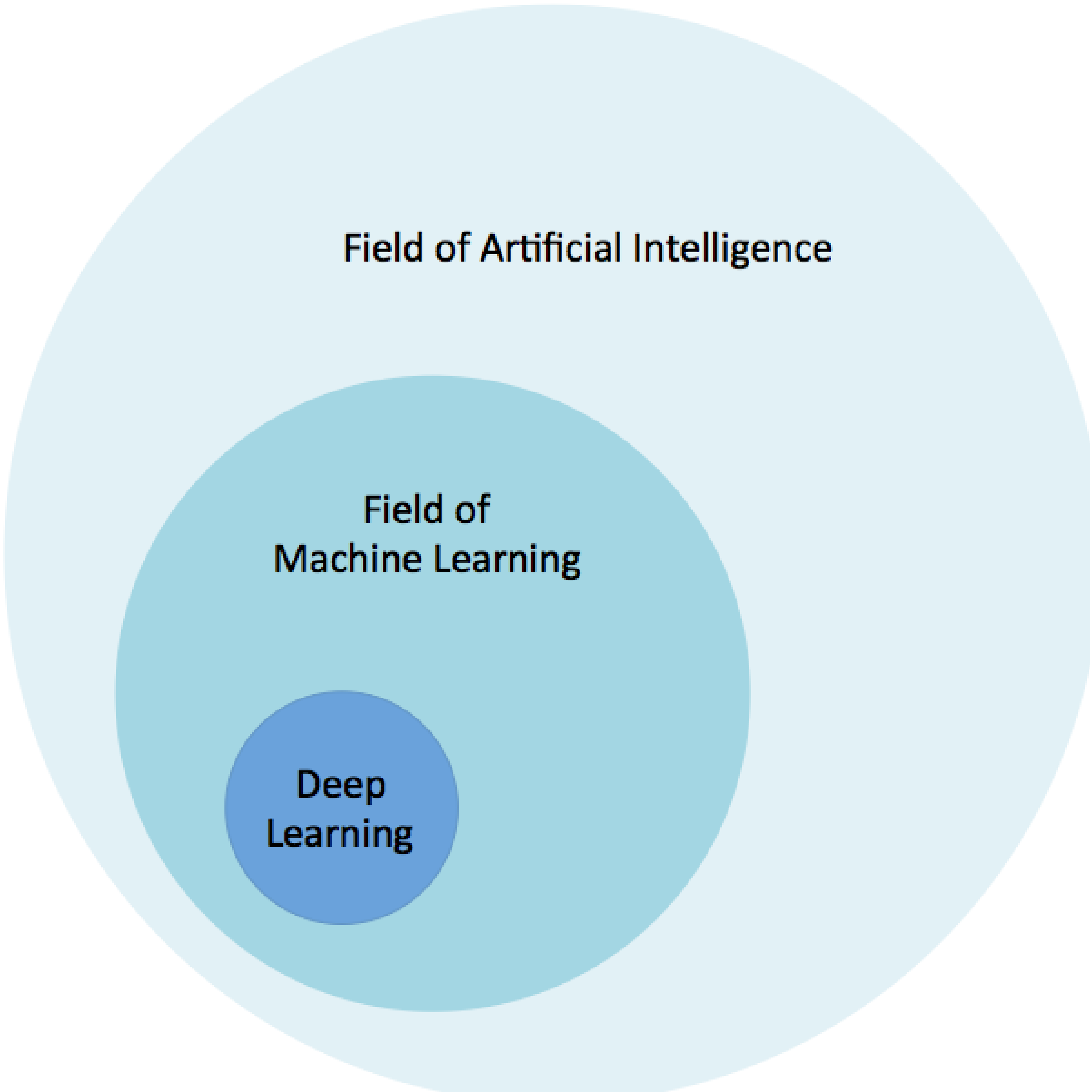
What is Machine Learning?

- Extracting Knowledge from raw data in the form of a model
  - Decision trees
  - Linear Models
  - Neural Networks
- Arthur Samuel quote:
  - "Field of study that gives computers the ability to learn without being explicitly programmed"

Why Machine Learning now?

- Computing resources are getting cheaper and more powerful
- More data is available

A Diagram


Field of Artificial Intelligence
Field of
Machine Learning
Deep
Learning

Machine Learning Compared to Data Science/Mining

- Data Mining
  - The process of extracting information from the data
  - Uses Machine Learning
- Data Science
  - Data Mining from the lens of a statistician
  - Venn Diagrams
  - A way to get a raise
  - A more agreeable Actuary
  - A statistician using a Mac

skymind

Machine Learning Examples

- Clustering
- Typically unsupervised learning
  - “K-Means Clustering”
- Example
  - “cluster K groups of similar news articles together”
- ND4J supports this, but it is not a NN

Machine Learning Examples continued...

- Random Decision Forest
- Recommender Engines
- Bayesian Classification

skymind

A Definition of Deep Learning

- Deep learning (also known as deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data.

source - wikipedia

skymind

Neural Networks

- A computational approach patterned on the human brain and nervous system

Comparison Between Neural Network and Machine Learning

- Machine Learning
    - Hand Crafted Features
    - SME(Subject Matter Expert) is needed
    - Must inject Context
- Deep Learning/Neural Network
    - Automatic Feature Engineering
    - Learns Context

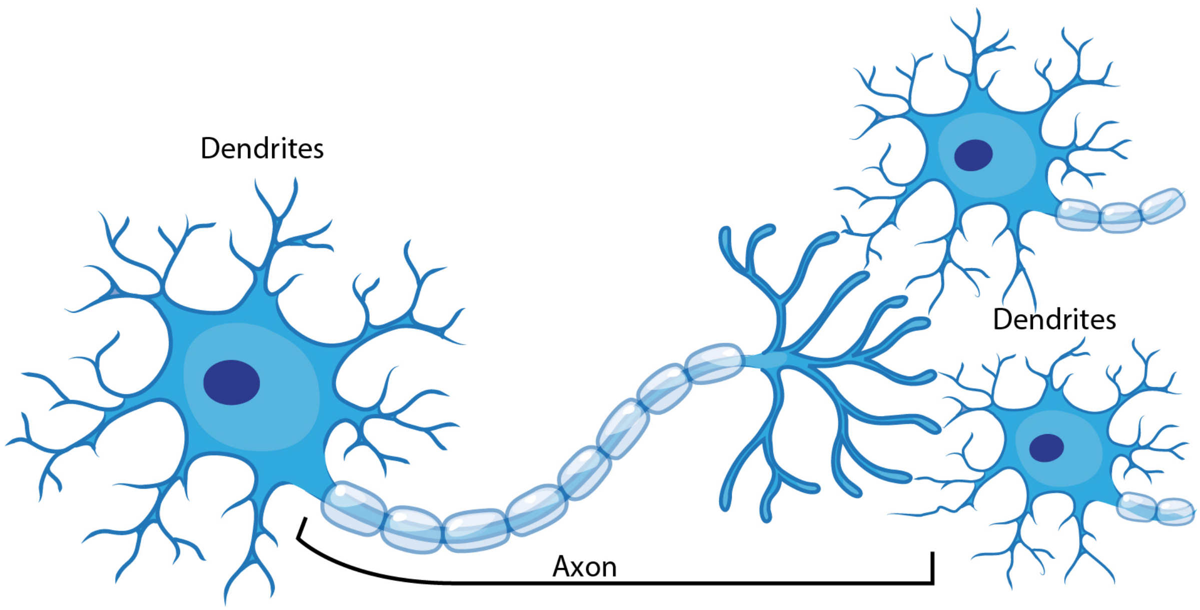
skymind

Biological Neurons

- Biological Neuron: An electrically excitable cell that processes and transmits information through electrical and chemical signals
- Biological Neural Network: An interconnected group of neurons

skymind

Biological Neuron



skymind

Role of Artificial Neural Network

Learns or Trains to perform tasks that traditional programming methods find rather challenging.

- Speech recognition
- object recognition
- computer vision
- pattern recognition.

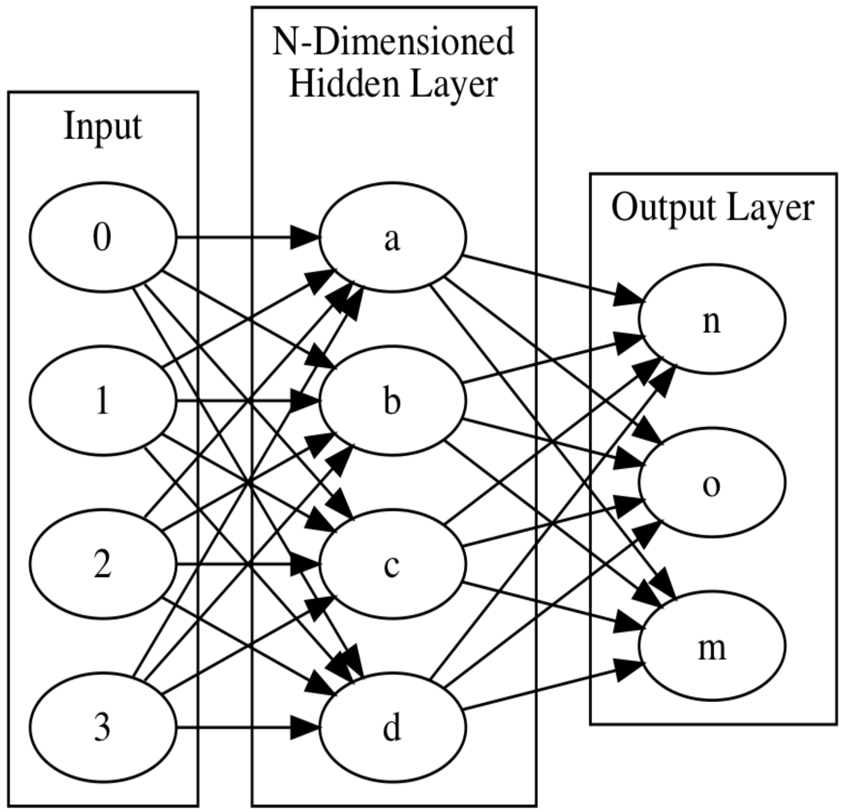
skymind

Supervised vs Unsupervised Learning

- Supervised learning
  - We give the training process labels ("outputs") for every training input data row
  - Model learns to associate input data with output value
- Unsupervised learning
  - No labels
  - Model attempts to learn structure in the data
- Neural Networks can be used for either supervised or unsupervised learning

skymind

A Neural Network



skymind

DeepLearning Considerations

- Inspired by the brain
    - Very basic implementation
- Brain has huge number of Neurons
- Brain has non-linear connections
- Creates similar distributed units of functionality
- Knowledge comes through connections

skymind

The Rise and Fall of Neural Networks

- 40 year old theory and practice
- Hype cycle followed by delusion
- Repeat cycle
- Incremental improvements over time

skymind

Why Neural Networks Now?

- 2012 Neural Network dominates image recognition
- Leads to current boom

Framing the Question

Using Neural Networks

Framing the Questions

- To build models we have to define
  - What is our training data (“evidence”)?
  - What kind of model (“hypothesis”) is appropriate for this data?
  - What kind of answer (“inference”) would we like to get from the model?
- These questions frame all machine learning workflows

What Neural Networks Do

- A = Input
- B = Expected Output
- Map A ==> B
- Using complex derivitable computation graph
    - Apply random weights at each edge
    - Adjust weights towards least error
    - repeat

skymind

A==> B Visually

	(Training Records) A					(Parameter Vector) x		(Actual Outcome) b
Input Record 1	0.7500000000000001	0.4166666666666663	0.702127659574468	0.5652173913043479		?		1.0
Input Record 2	0.6666666666666666	0.5	0.9148936170212765	0.6956521739130436	•	?	=	2.0
Input Record 3	0.4583333333333326	0.3333333333333336	0.8085106382978723	0.7391304347826088		?		2.0

Linear Algebra Terms

- Scalars
  - Elements in a vector
  - In compsci synonymous with the term “variable”
- Vectors
  - For a positive integer n, a vector is an n-tuple, ordered (multi)set, or array of n numbers, called elements or scalars
- Matricies
  - Group of vectors that have the same dimension (number of columns)

Linear Algebra Terms Continued..

- Scalar as point
  - one dimension
- Vector as line
  - two dimensions
- matrix as plane
  - three dimensions
- Tensor
  - More than 3 dimensions
  - Tensor == NDarray

Everything is a Tensor

- Rank of NDArray == Number of Dimension
- Rank 0 == scalar
- Rank 1 == vector
- Rank 2 == matrix

skymind

Solving Systems of Equations

- Two general methods
  - Direct method
  - Iterative methods
- Direct method
  - Fixed set of computation gives answer
  - Data fits in memory
  - Ex: Gaussian Elimination, Normal Equations
- Iterative methods
  - Converges after a series of steps
  - Stochastic Gradient Descent (SGD)

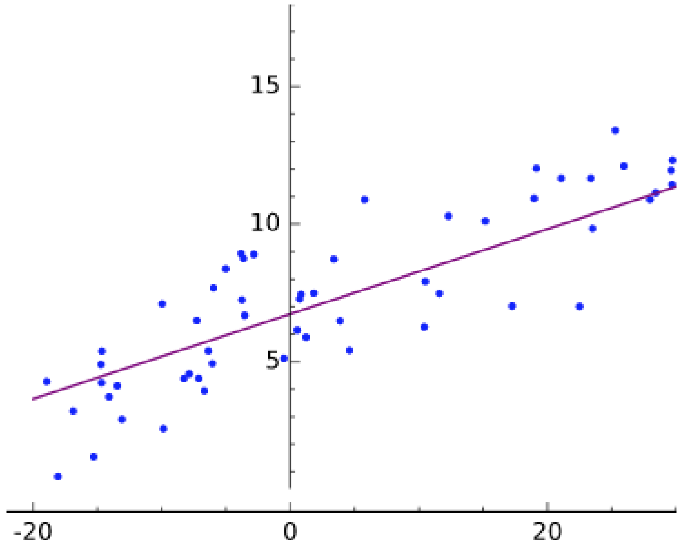
Neural Networks Use an Iterative Method to Solve a System of Equations

- Values are tried
- Error is calculated
- Values are updated

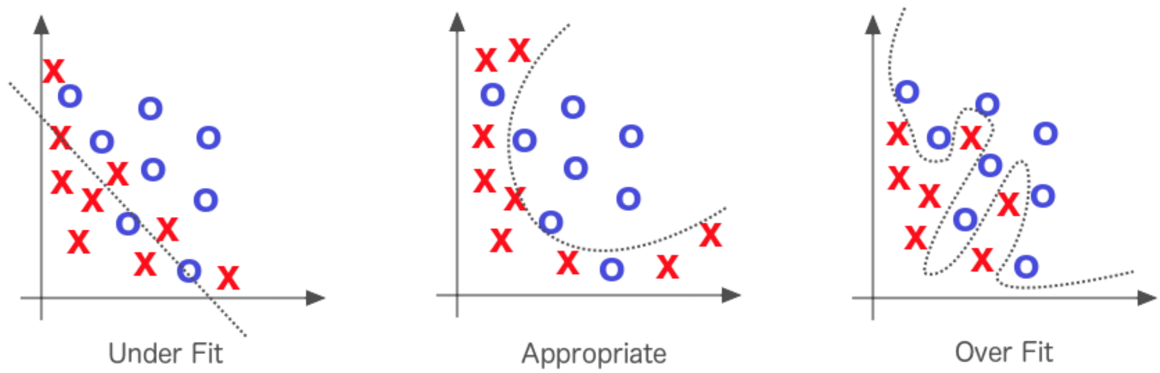
Training a Neural Net

- Inputs: Data you want to produce information from
- Connection weights and biases govern the activity of the network
- Learning algorithm changes weights and biases with each learning pass

Fitting the Training Data



Overfitting the Training Data

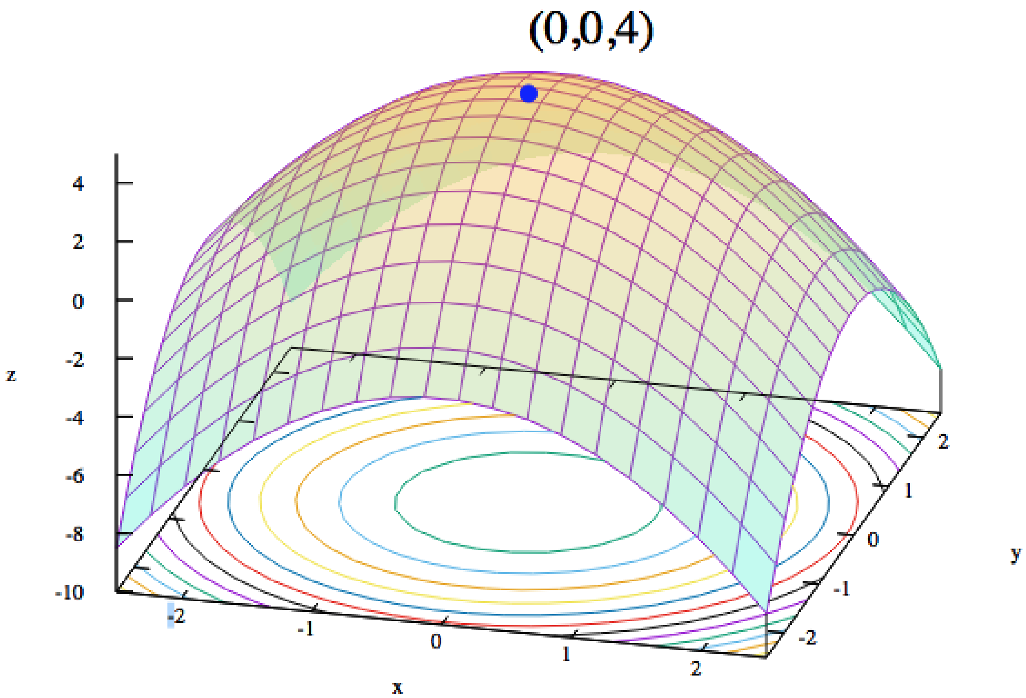


Optimization

- Iteratively adjust the values of the x parameter vector
    - Until we minimize the error in the model
- Error = prediction – actual
- Loss functions measure error
    - simple/common loss function:
    - “mean squared error”
- How do we make choices about the next iterative “step”?
    - Where “step” is how we change the x parameter vector

Loss Function

- Assigns cost to output vs expected output
- Optimization seeks to minimize the value of the loss function


(0,0,4)
z
4
2
0
-2
-4
-6
-8
-10
x
-2
-1
0
1
2
y
2
1
0
-1
-2

Gradient Descent

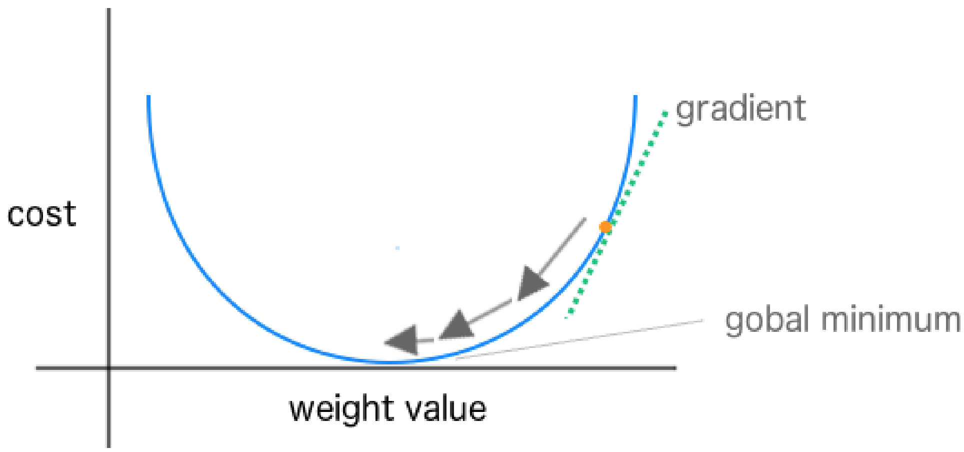
- Optimization method where we consider parameter space as
  - “hills of error”
  - Bottom of the loss curve is the most “accurate” spot for our parameter vector
- We start at one point on the curved error surface
  - Then compute a next step based on local information
- Typically we want to search in a downhill direction
  - So we compute the gradient
    - The derivative of the point in error-space
    - Gives us the slope of the curve

Stochastic Gradient Descent

- With basic Gradient Descent we look at every training instance before computing a “next step”
- With SGD with compute a next step after every training instance
  - Sometimes we’ll do a mini-batch of instances

skymind

SGD Visually Explained



Summary

- A Neural Network Transforms input to output through a process of
    - Computation Graph of complex non-linear functions
    - Random weight initialization
    - Update of weights after calculating loss function to improve results
    - Iterate to further improve results

skymind