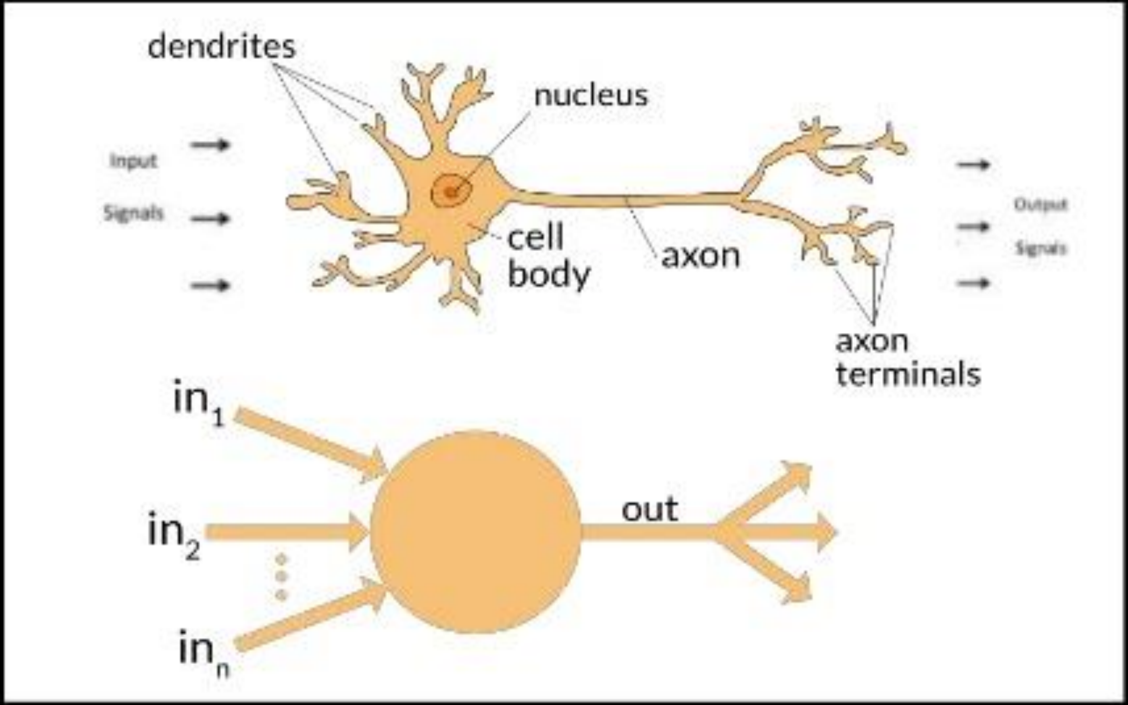
IST407/707 Applied Machine Learning

Neural Networks and Deep Learning

Syracuse University
School of Information Studies

What is a Neural Network?

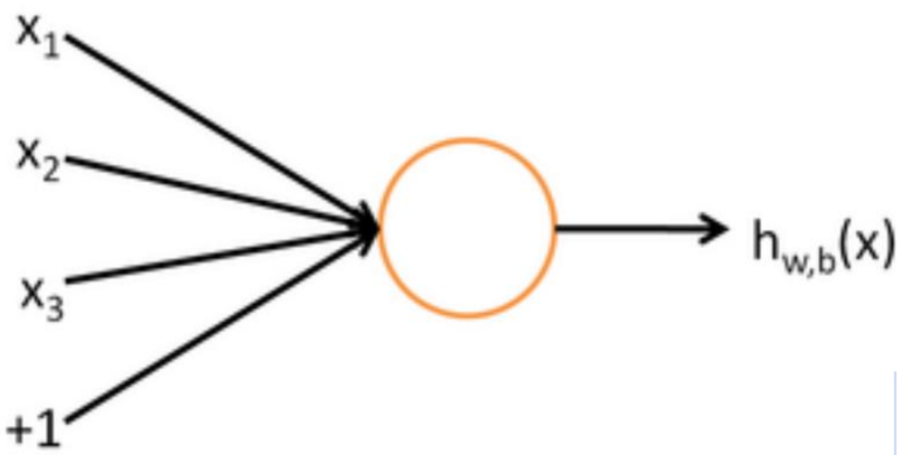
- Intuition: create an Artificial Neural Network to solve problems in the same way as the human brain



Single Neuron

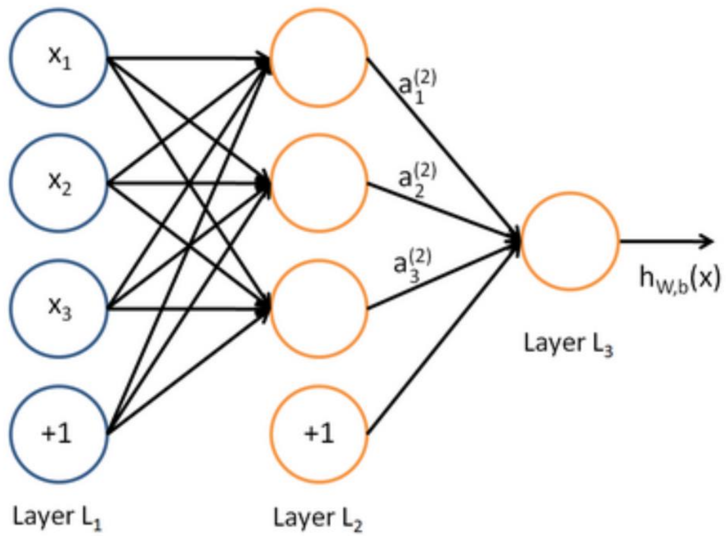
A single neuron is a computational unit with an activation function (f). It takes inputs (3) plus a bias term (b) and gives an output

- Expressed as a result h depending on weights w and b



Single Layer Neural Network (NN)

- Can also solve current machine learning problems
- If we feed a vector of inputs (the features) into a number of neurons, we get a vector of outputs, which can be combined according to the task we are trying to solve (the objective function h)



Input Layer Hidden Layer Output Layer

Training the network learns the weights:

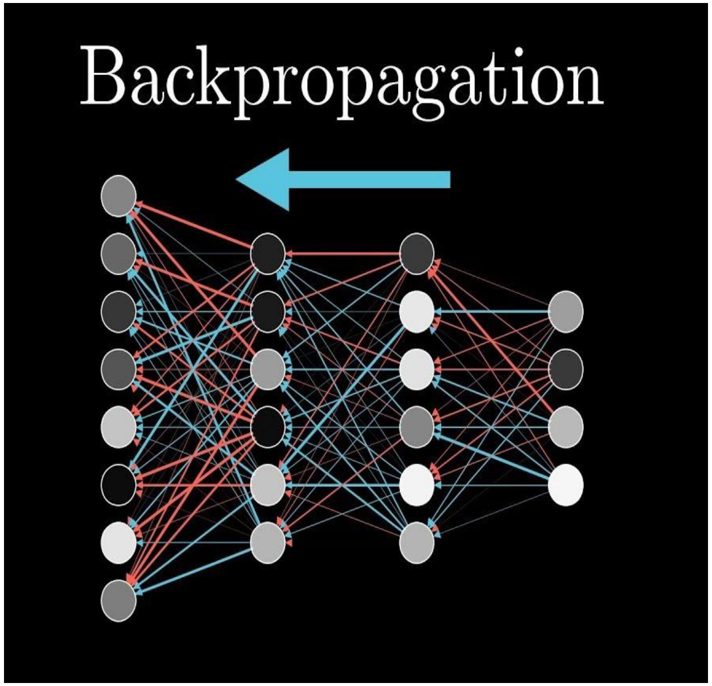
- run the network to predict an output,
- compare the output with the desired (gold) result,
- run back through the network adjusting the weights to reduce the error,
- and iterate.

The weights (and bias) gives the model to compute the predicted output for future data.

Backpropagation

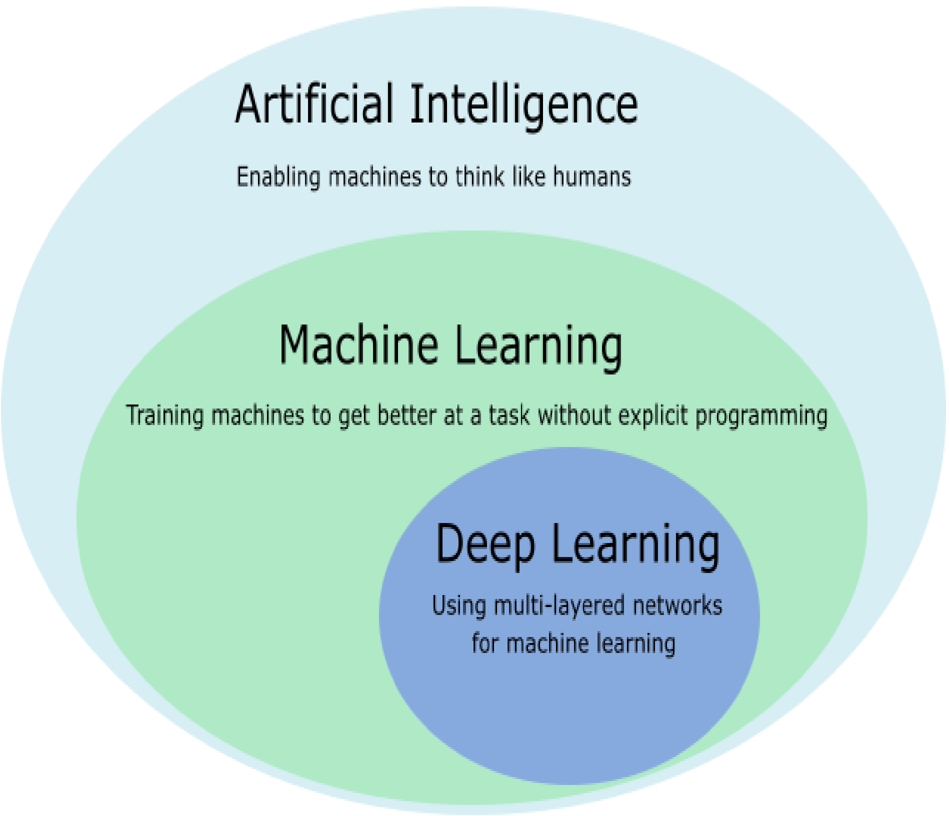
The core algorithm for setting the weights of a neural network

- *Based on the error rate (i.e. loss)* obtained in the previous epoch (i.e. iteration).
- Proper tuning of the weights ensures lower error rate
- Makes the model reliable by increasing its generalization.



[Reference: https://www.youtube.com/watch?v=Ilg3gGewQ5U](https://www.youtube.com/watch?v=Ilg3gGewQ5U)

What is deep learning？



- An artificial intelligence capability that uses machine learning to imitate the human brain
- Creates a network of “neurons” capable of learning from data
- Also known as:
  - Deep neural learning
  - Deep neural network.

https://www.datamation.com/artificial-intelligence/ai-vs-machine-learning-vs-deep-learning.html

Deep Learning Architecture

Most commonly used architecture uses various types of multi-layer neural networks, such as Belief NN

Output layer

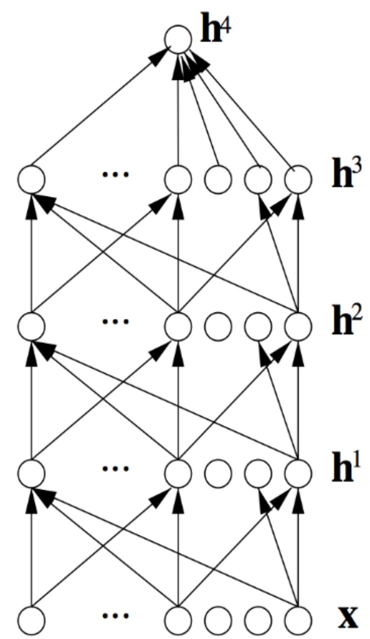
Here predicting a supervised target

Hidden layers

These learn more abstract representations as you head up

Input layer

Raw sensory inputs (roughly)

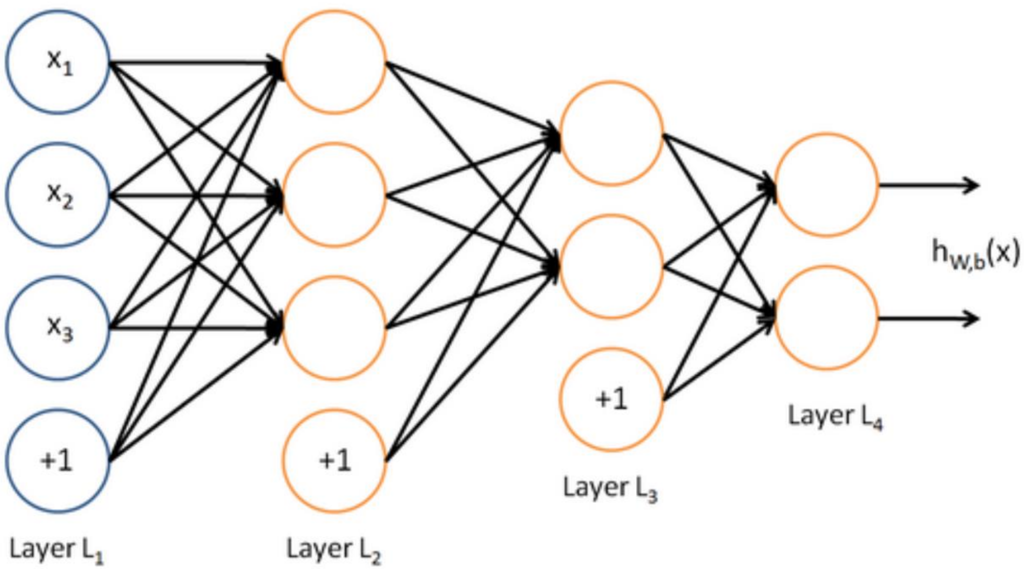


Deep Learning NN

We can keep going and add multiple layers

And we can revise our learning algorithm to also learn representations of the input X

- Several algorithms for how to do the "feed forward" and "back propagation" in an efficient way



Why multiple layers of Neurons?

A *single-layer* neural network can only be used to represent *linearly separable* functions:

- Problems such as when the two classes in a classification problem can be neatly separated by a line

However, most problems are not linearly separable

A *multilayer* network can be used to represent *convex* regions:

- The networks can learn to draw shapes around examples in a high-dimensional space that can separate and classify the regions
- This overcomes the limitation of linear separability

Deep Learning vs SVMs

SVMs are non-parametric models:

- **Complexity grows as the training size increases**
  (i.e. training the model can be expensive computationally)

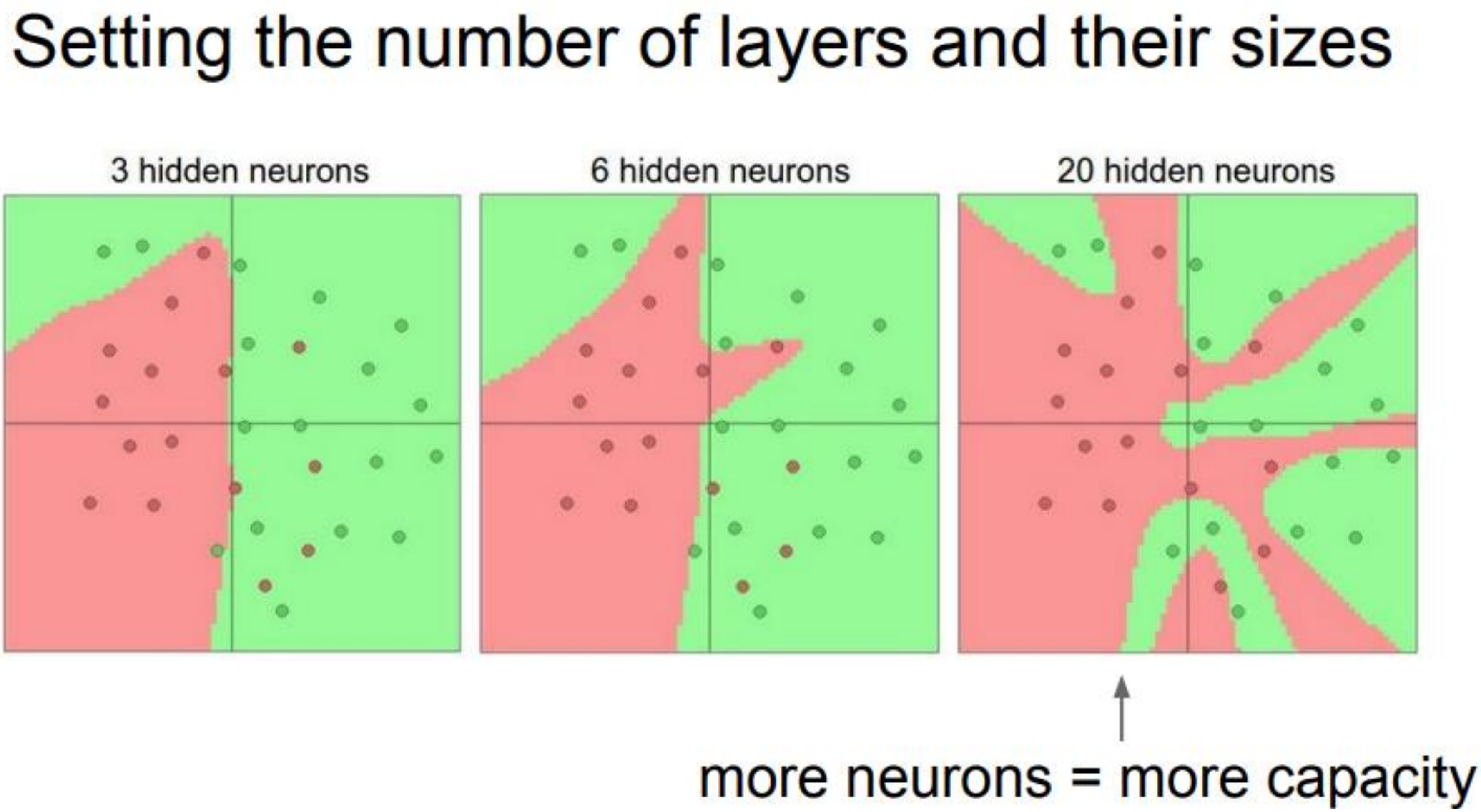
- In a complex dataset worst-case scenario, we can end up with as many support vectors as we have samples in the training set.

SVMs and other simpler models are typically preferred for relatively small data sets with fewer outliers.

Deep learning needs relatively large datasets (and infrastructure) to train the model.

Raschka, S. (2016, April 22). When Does Deep Learning Work Better Than SVMs or Random Forests? Retrieved from https://www.kdnuggets.com/2016/04/deep-learning-vs-svm-random-forest.html

The Impact of More Neurons



http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture04.pdf

Deep Learning Example

Tinker With a Neural Network Right Here in Your Browser.
Don’t Worry, You Can’t Break It. We Promise.

[Try it out]

Basic Deep Learning Vocabulary

Epoch: an iteration of learning (one pass of the training data)

Accuracy: % of correct predictions (e.g., based on confusion matrix)

Loss (Loss Function): how inaccurate is the predictions by looking at prediction % (predicting 90% for the correct choice is better than predicting 70%)

Tensor: a generalization of vectors and matrices
(i.e., a multidimensional array)

Two main parameters control the architecture (or topology) of the Deep Learning network:

- The number of **layers** in the network
- The number of **neurons (nodes)** in each layer

Additional Information

Understanding how backpropagation trains:

https://google-developers.appspot.com/machine-learning/crash-course/backprop-scroll/

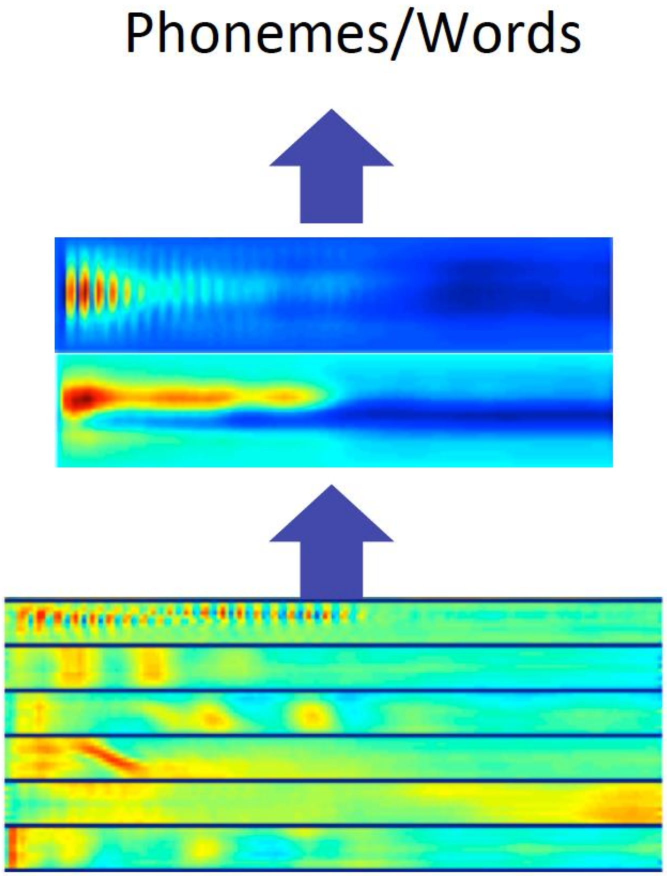
Exploring Deep Neural Networks (Tensor Flow playground):

https://cloud.google.com/blog/products/gcp/understanding-neural-networks-with-tensorflow-playground

DEEP LEARNING AND NLP

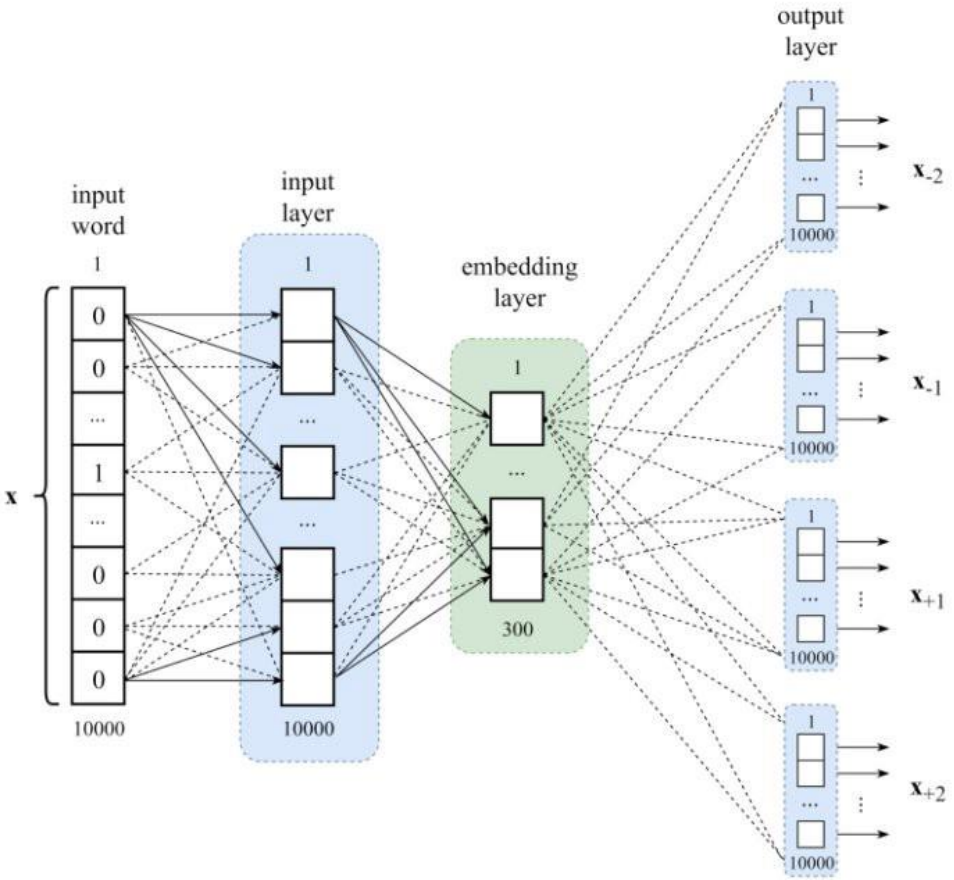
Deep Learning for Speech

- The first breakthrough results of deep learning on large datasets happened in speech recognition
- Context-dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition (Dahl et al 2010)



Deep learning word embeddings

- Self-supervised; uses attempt to “fix” noise in the input layer
- The internal representation of each input word is a pattern of activation in the hidden layer
- This is called the “embedding” layer
- The embeddings for each word form a vector space that can be used in subsequent tasks

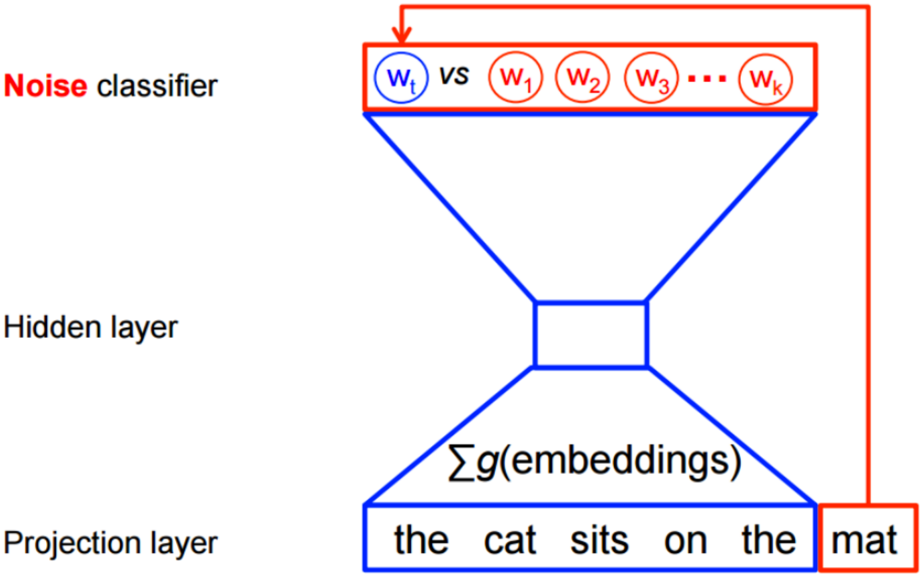


Word Embedding Example

Collobert et al JMLR 2011 : Set up a classification task from unsupervised data where we have positive training examples directly from the data, and negative examples obtained by substituting a random word in the context

- Positive example:
  "cat sits on the mat"
- Negative example:
  "cat sits jeju the mat"

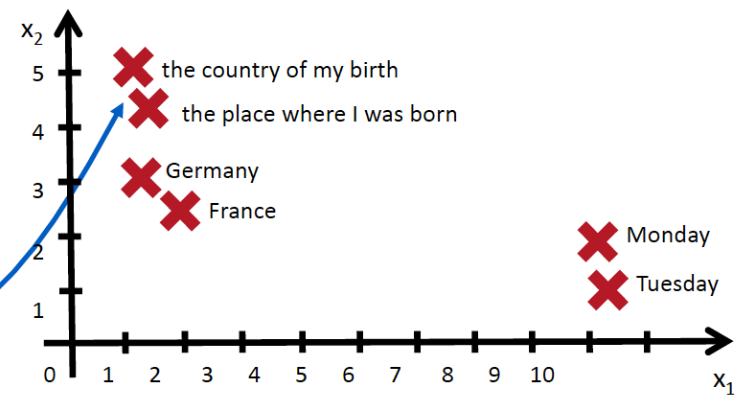
Classify which contexts are noise

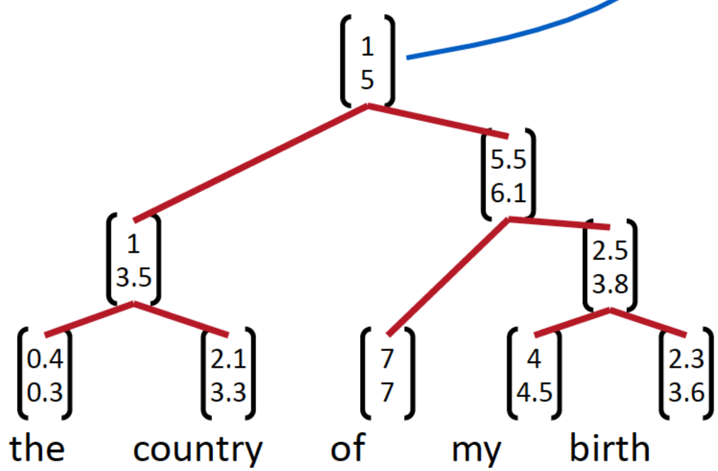


Phrase Level Vectors

Represent the meaning of longer phrases by mapping them into the same vector space

Use principle of compositionality
The meaning (vector) of a sentence is determined by (1) the meanings of its words and (2) the rules that combine them.

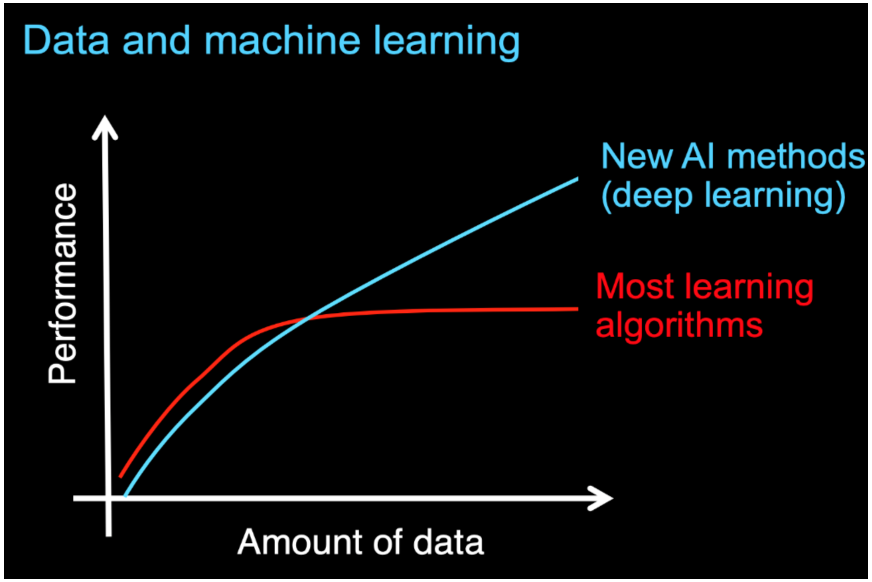




Models in this section can jointly learn parse trees and compositional vector representations

DEEP LEARNING IN THE REAL WORLD

Why is deep learning useful?



Deep learning:

- Utilizes large amounts of training data
- Provides a **flexible**, framework for representing world, visual and linguistic information.

Feature Engineering:

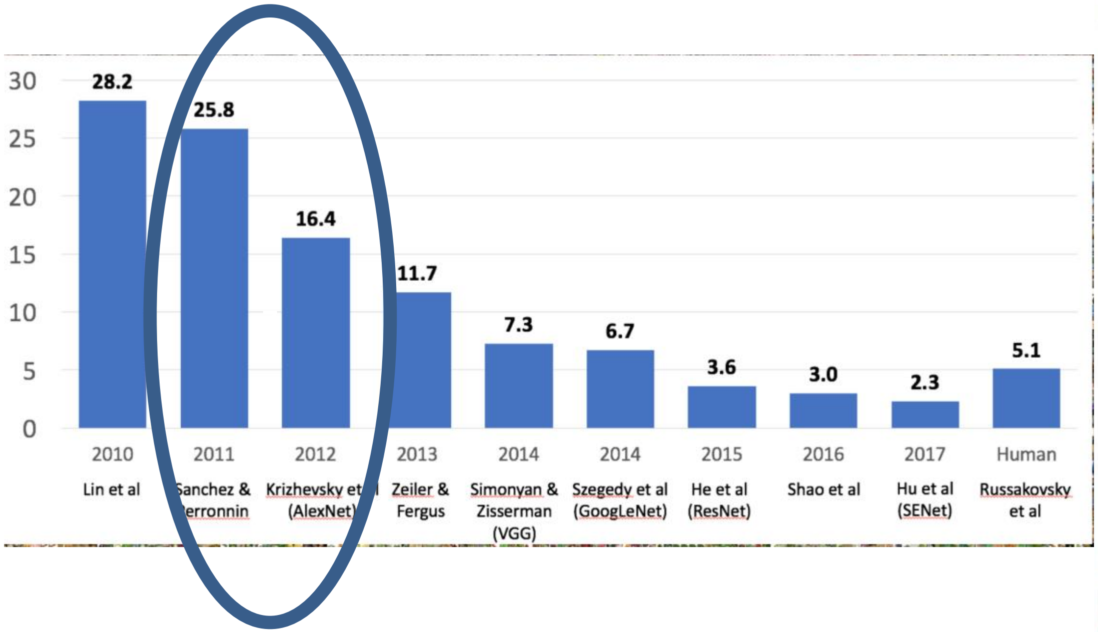
- Manually designed features are often **over-specified**, **incomplete** and take a **long time to design** and validate
- Machine Learned Features are **easy to adapt**, **fast** to learn

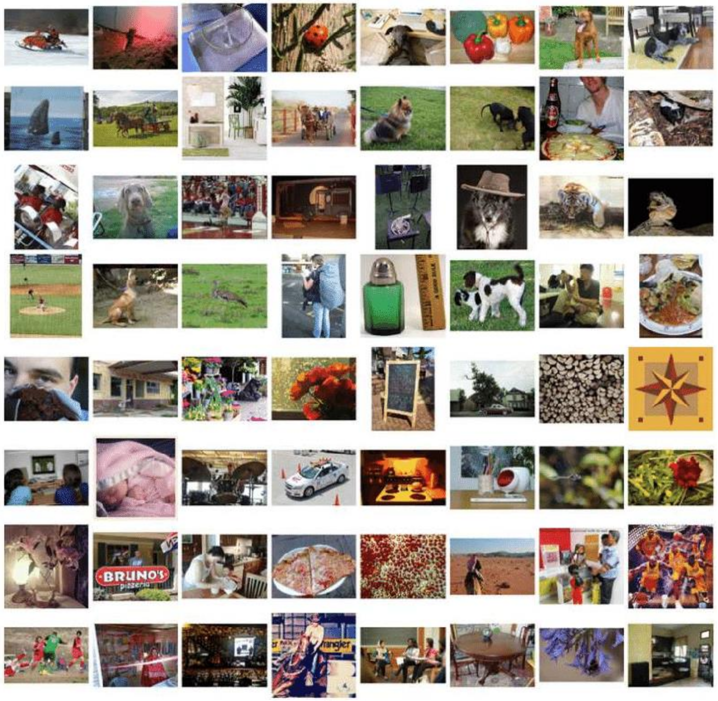
Reference
http://cs229.stanford.edu/materials/CS229-DeepLearning.pdf

Deep Learning Impact: Image Recognition

ImageNet Large Scale Visual Recognition Challenge:

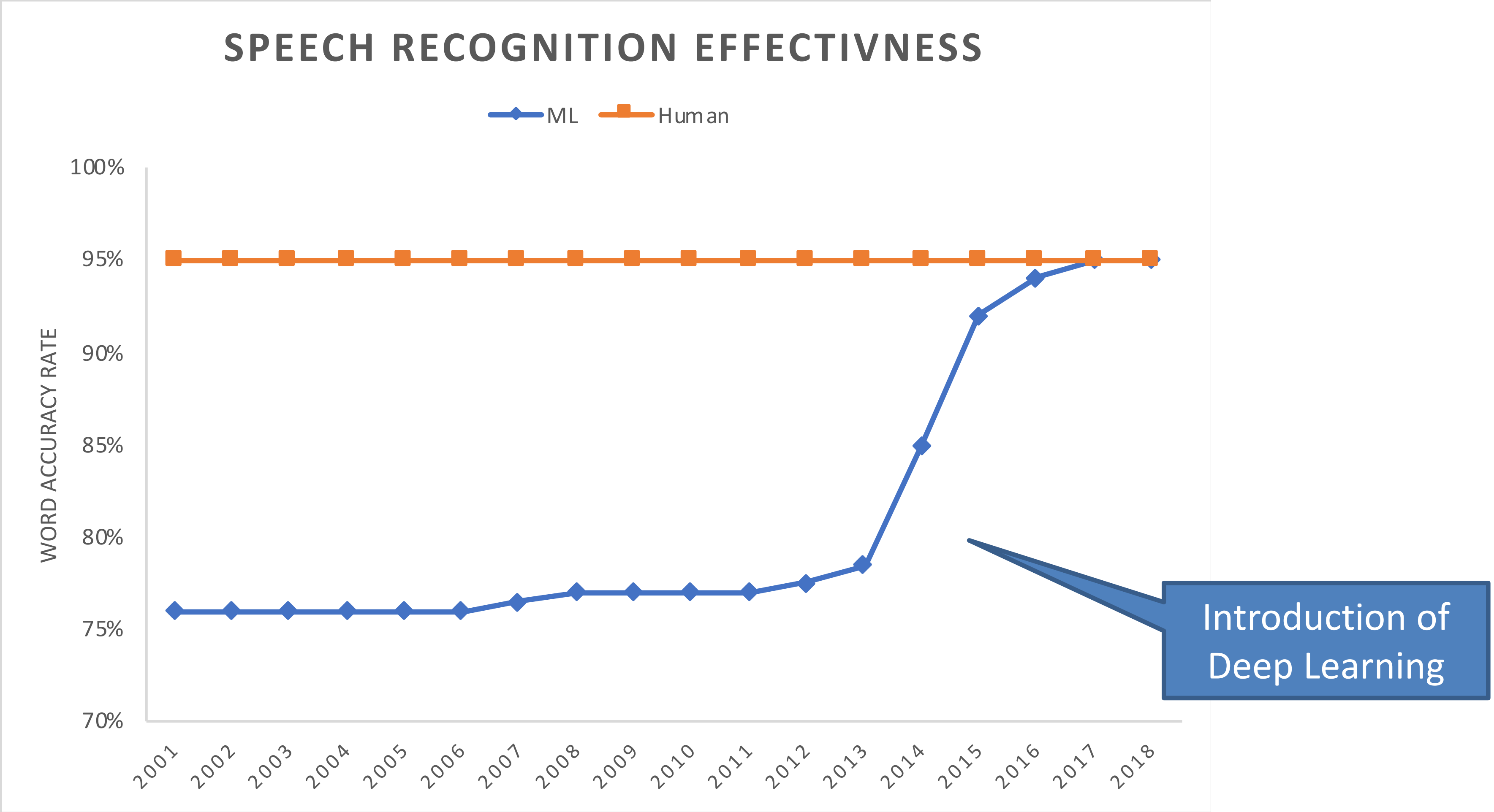
- When deep learning was first used (2012), the deep neural network was 41% better than the previous best solution





https://www.researchgate.net/figure/Examples-in-the-ImageNet-dataset_fig7_314646236
Imagenet info: https://ruder.io/nlp-imagenet/
http://cs231n.stanford.edu/slides/2019/cs231n_2019_lecture01.pdf

Deep Learning Impact: Speech recognition



https://www.slideshare.net/kleinerperkins/internet-trends-2017-report

TensorFlow

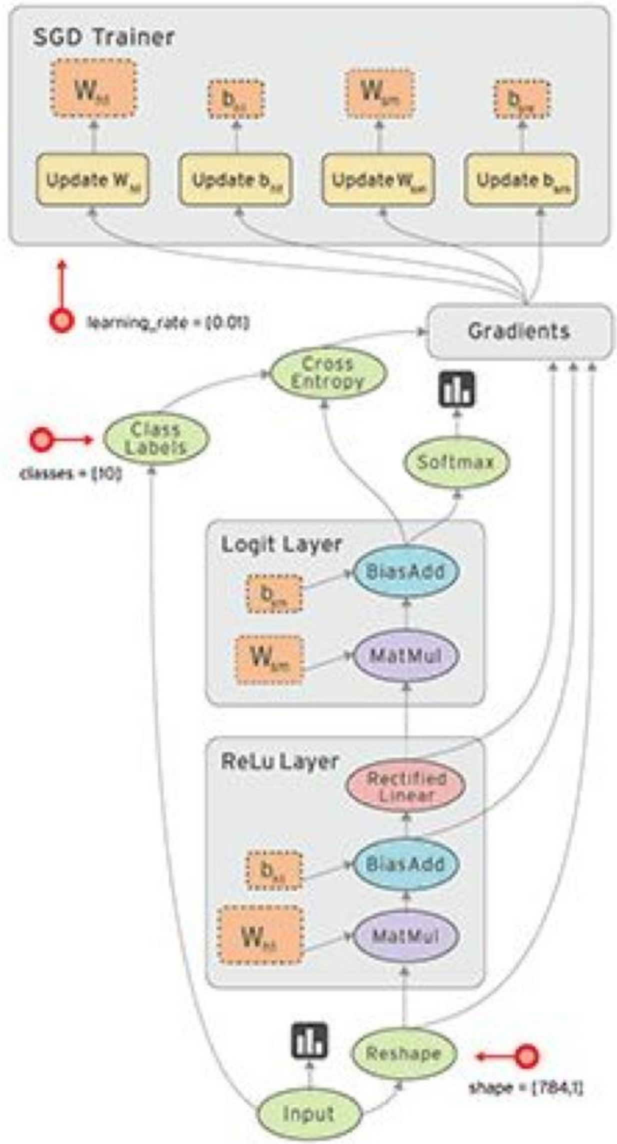
Framework from Google for scalable ML
- Framework of reusable components

Define computation as a graph
- Graph is defined with python fns
- Compiled, optimized, executed
- Nodes represent computations
- Data (tensors) flow along edges

Manage distributed, heterogeneous systems

But must program details of neural net classifiers

Higher level packages in progress
- E.g



Learning resources

Free online tutorial, blogs, and videos available

- LinkedIn Learning at Syracuse University (for SU students, faculty and Staff) (https://answers.syr.edu/display/lnkdlrn/LinkedIn+Learning+at+Syracuse+University)

- *10 Free Courses for Machine Learning and Data Science* (https://www.kdnuggets.com/2018/11/10-free-must-see-courses-machine-learning-data-science.html)

- *R Markdown Notebooks for "Deep Learning with R"* *(*https://github.com/jjallaire/deep-learning-with-r-notebooks*)*