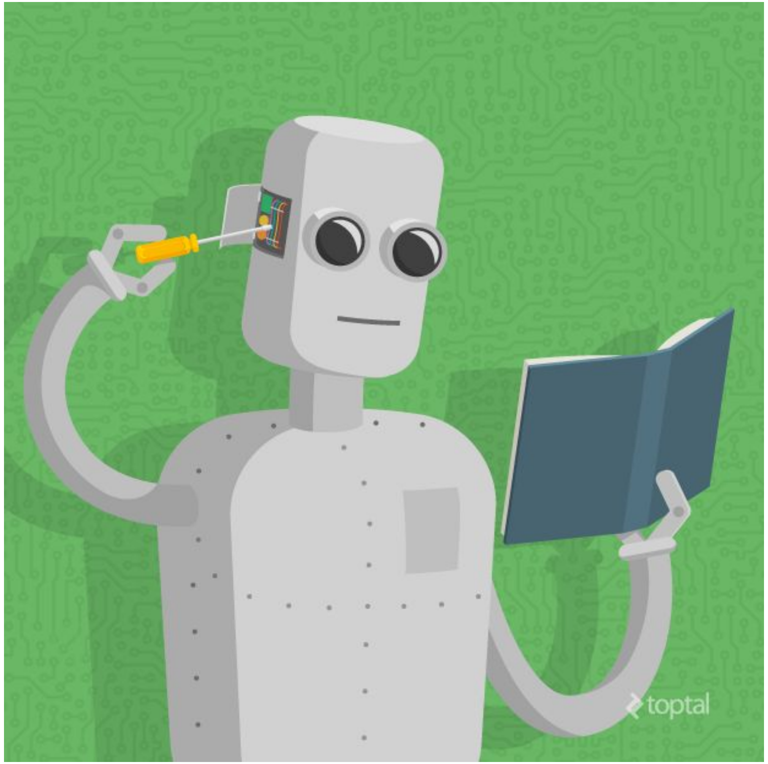
Masterclass: Introduction to Deep Learning

Kyle Swanson

Massachusetts Institute of Technology

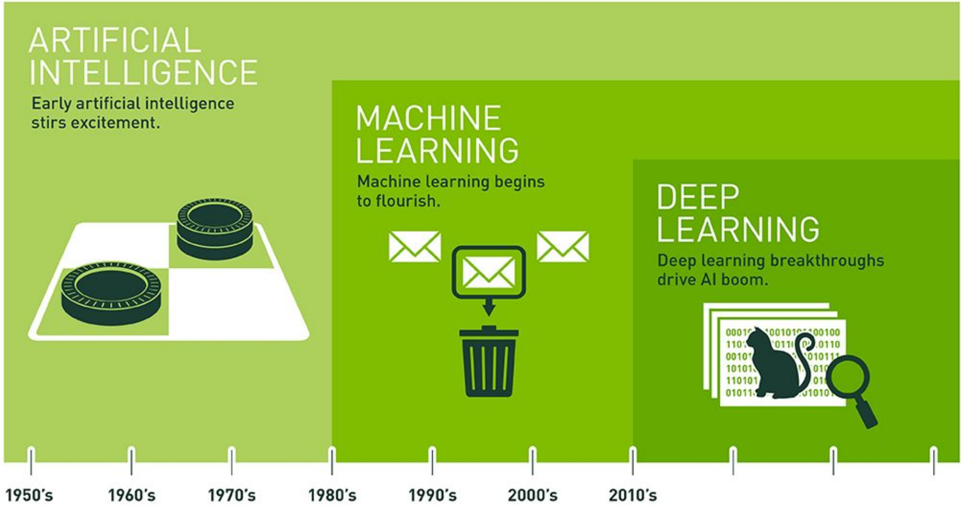
Today

- Machine learning recap
- What is deep learning?
- How do neural networks work?
- Deep learning applications
- Case study
    - Deep learning for breast cancer detection



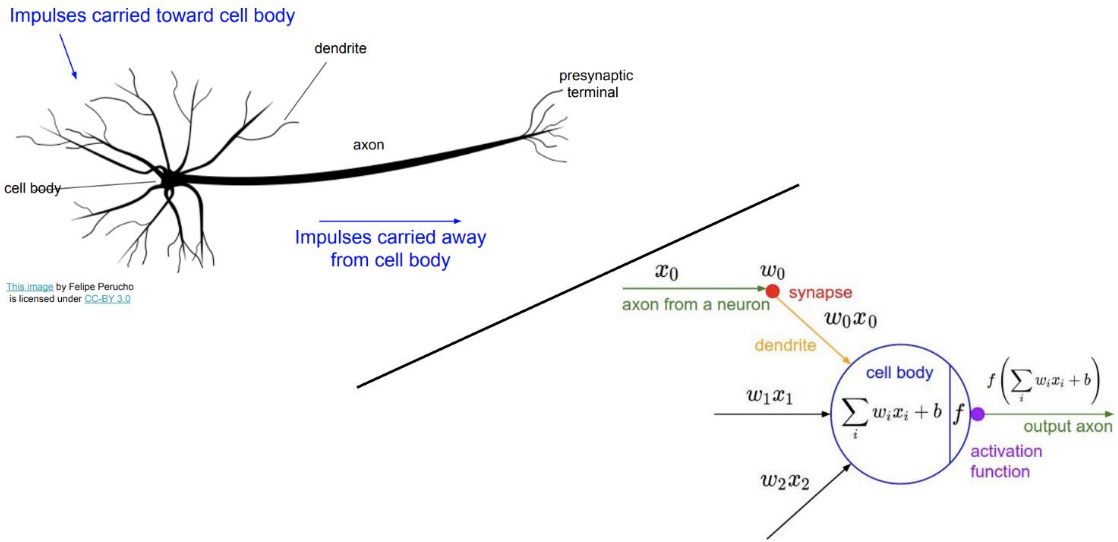
Machine learning recap

- **Machine learning:** “Field of study that gives computers the ability to learn without being explicitly programmed.” (Arthur Samuel, 1959)
- Types of problems
  - Classification
  - Regression
  - Generation
- Types of learning
  - Supervised learning
  - Reinforcement learning
  - Unsupervised learning
- Applications
  - Translation, chatbots, facial recognition, stock market prediction, movie recommendation, etc.

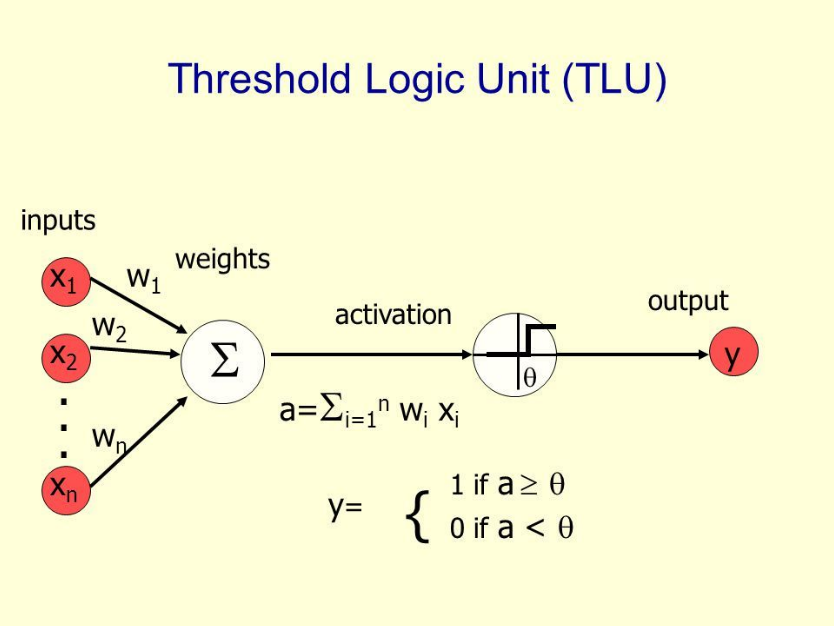


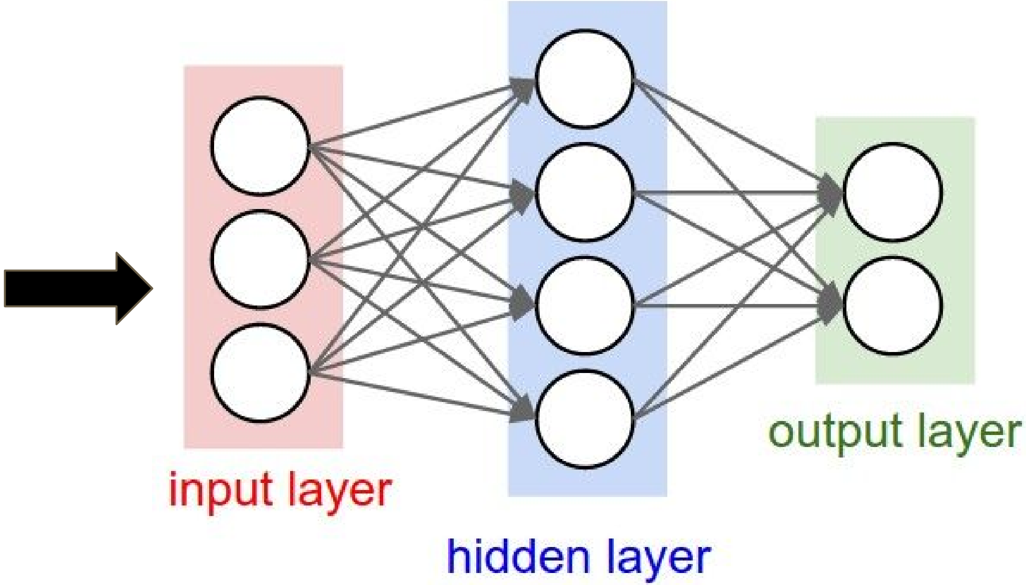
What is deep learning?

- **Deep learning:** A subfield of machine learning concerned with algorithms inspired by the structure and function of the brain.
- These algorithms are called **neural networks**



From one neuron to many

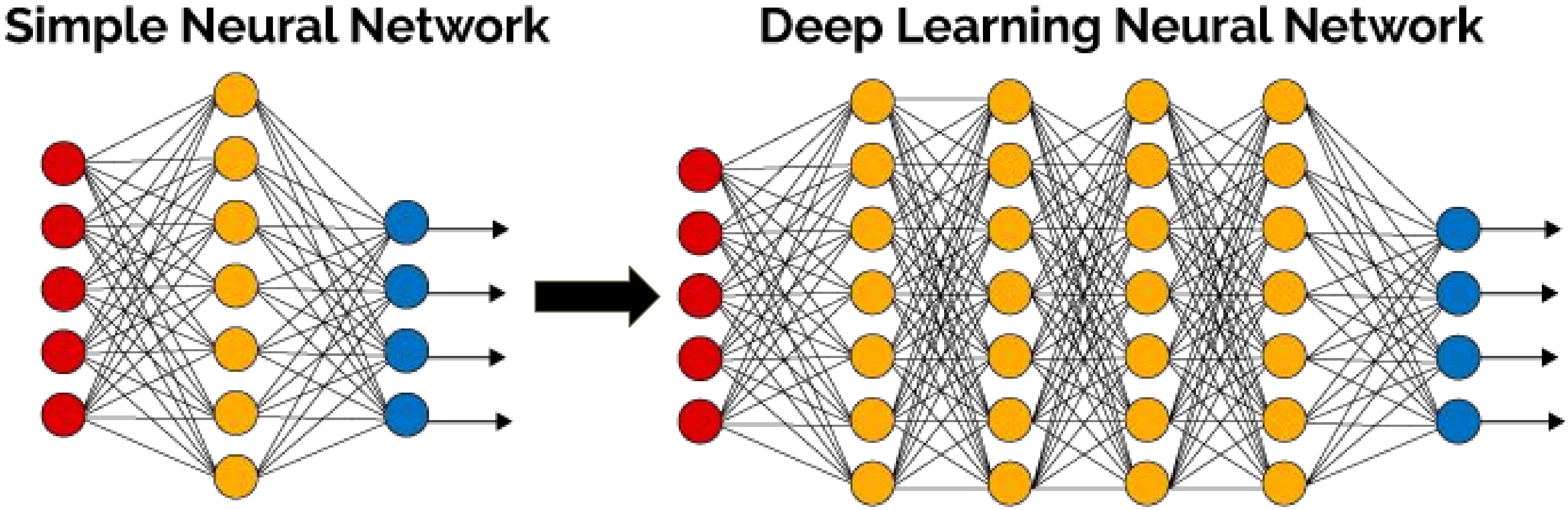




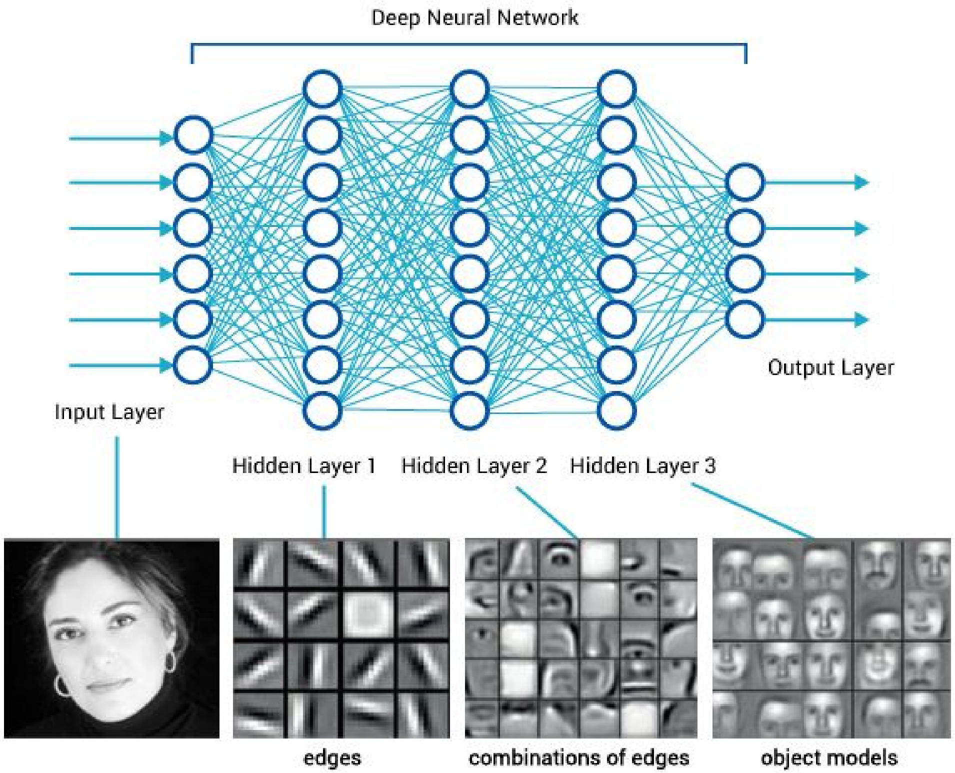
Universal approximation theorem

- **Math:** A single layer neural network with a finite number of neurons can approximate continuous functions on compact subsets of $\mathfrak{R}^n$
- **English:** A single layer neural network can approximate any reasonable function (i.e. it can compute virtually anything)
- **Reality:** Approximating complex functions with a single layer is hard
- **Solution:**

From one layer to many



Multiple layers build abstractions

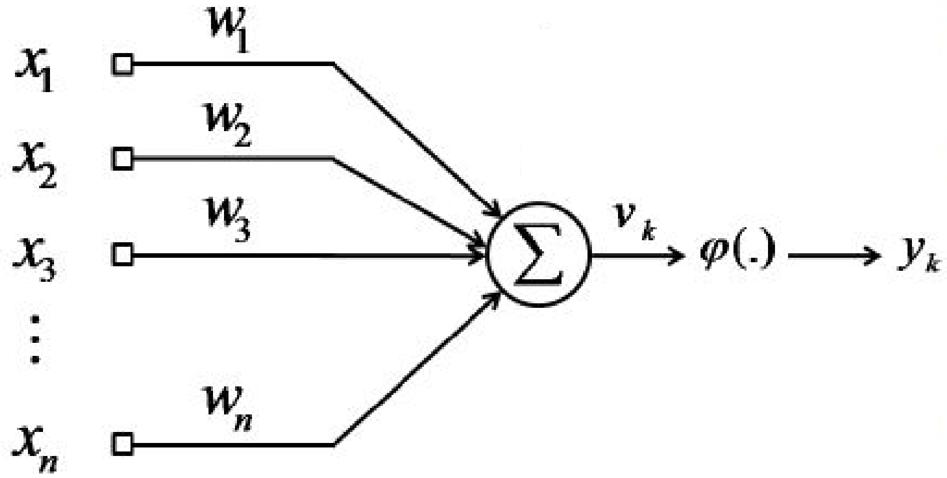


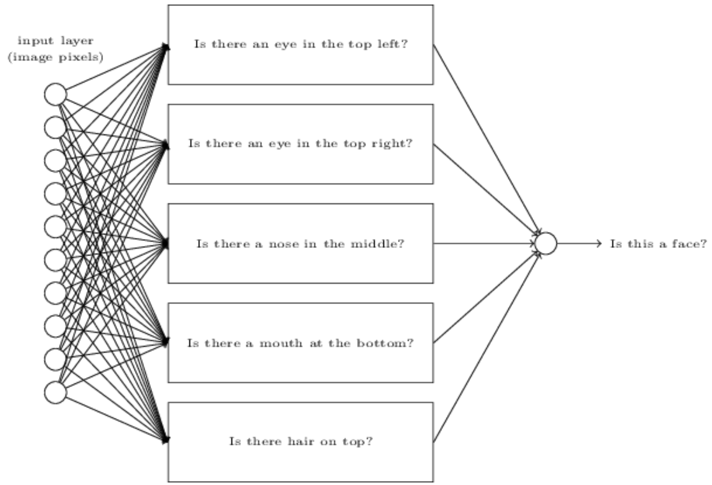
How do neural networks work? - neurons

- Each neuron computes a weighted sum of the inputs
- Different neurons have different weights, meaning neurons focus on different features of the input
- A non-linear function is applied to the weighted sum

$$v_k = \sum_{i=1}^{n} w_i x_i$$

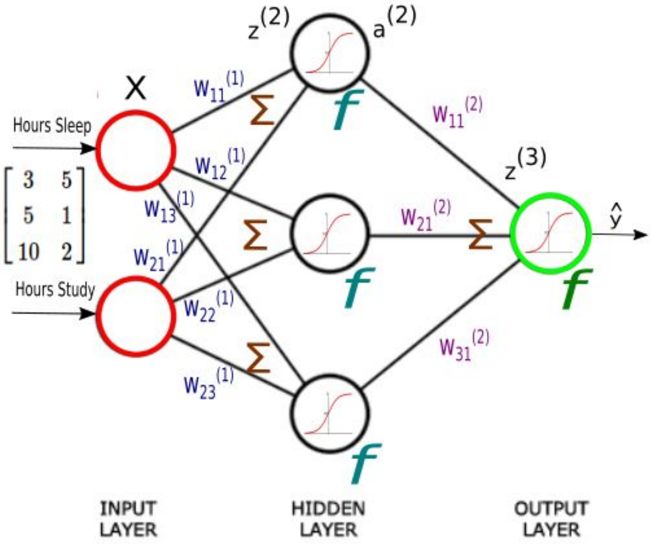
$$y_k = \phi(v_k)$$

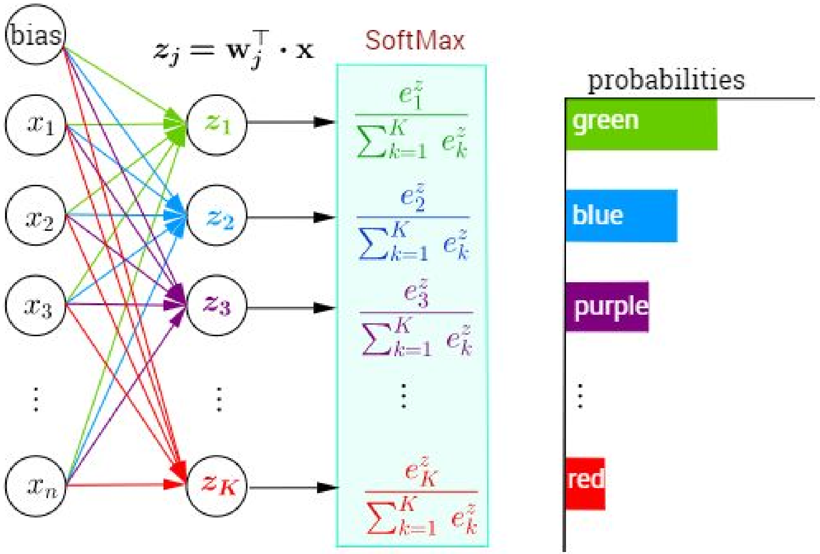




How do neural networks work? - prediction

- Output of one layer becomes input of the next layer
- Final prediction is the output of the last layer
    - Special non-linear function on last layer to output probability distribution (softmax)



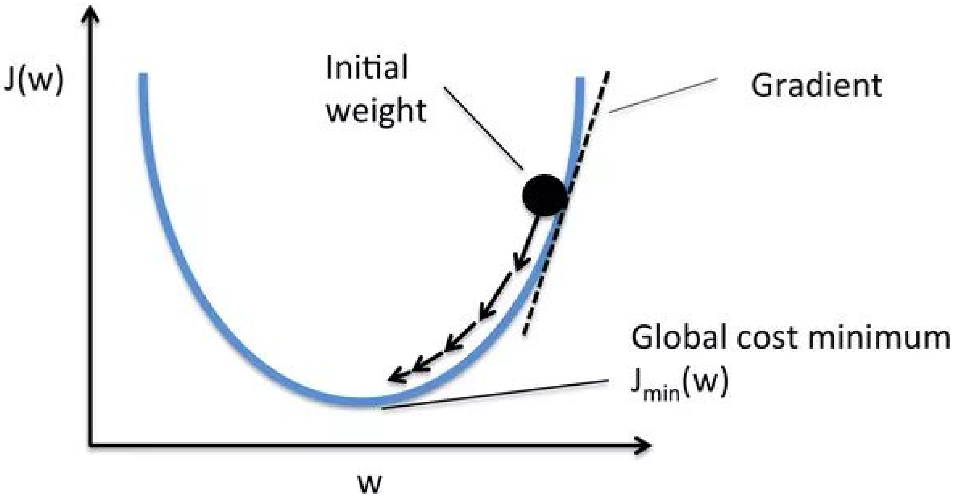


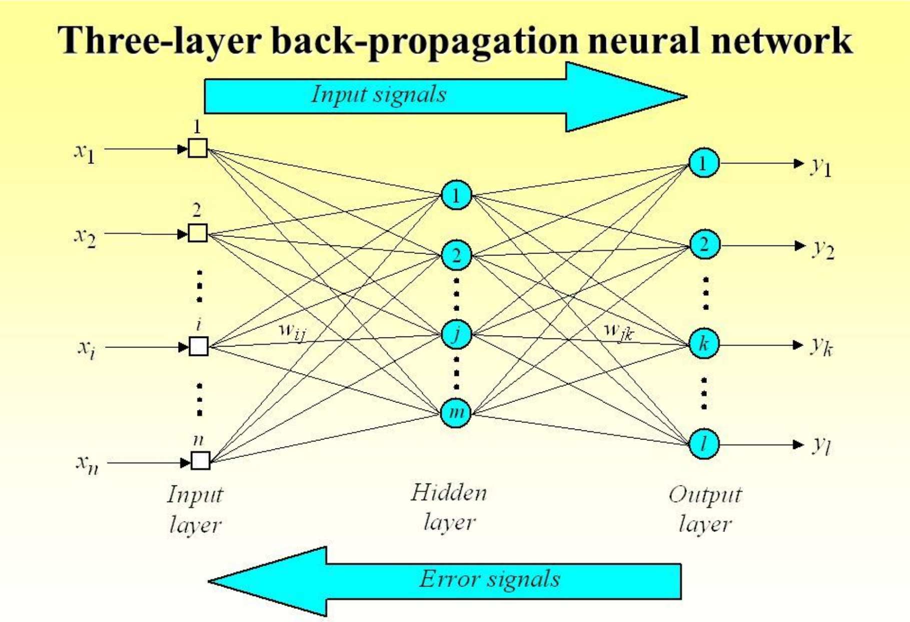
How do neural networks work? - learning

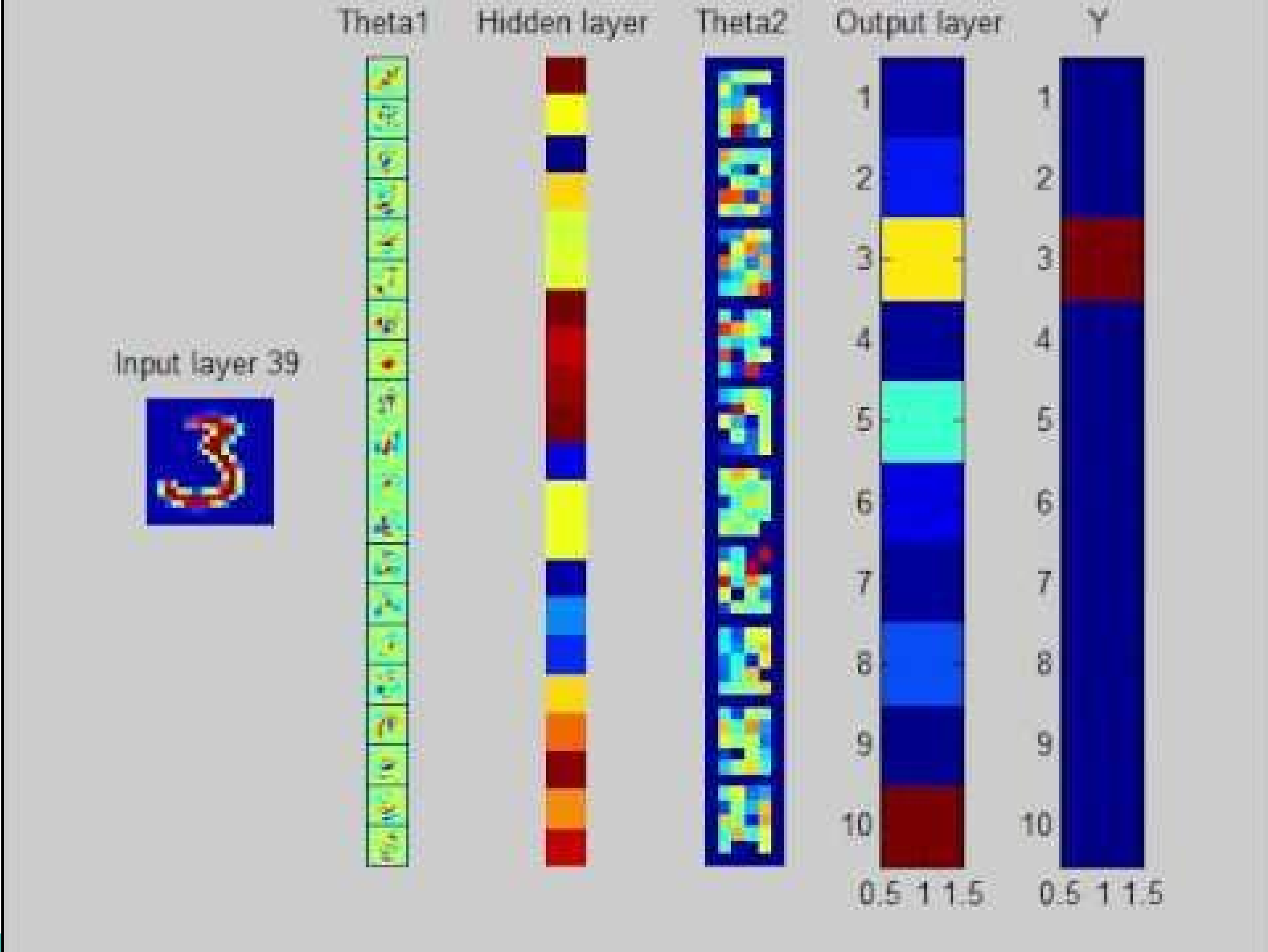
- Different weights lead to different predictions
- How do we learn the weights to make the correct prediction?
- Define a loss function - J(w)
  - A function of the weights of the network
  - Quantifies how far away we are from the correct prediction
  - Lower loss is better
- Backpropagation algorithm (calculus)
  - Determine how changing each weight changes the loss
    - Compute the gradient of the loss with respect to the weight
  - Change the weight to minimize the loss
    - Perform a gradient descent update

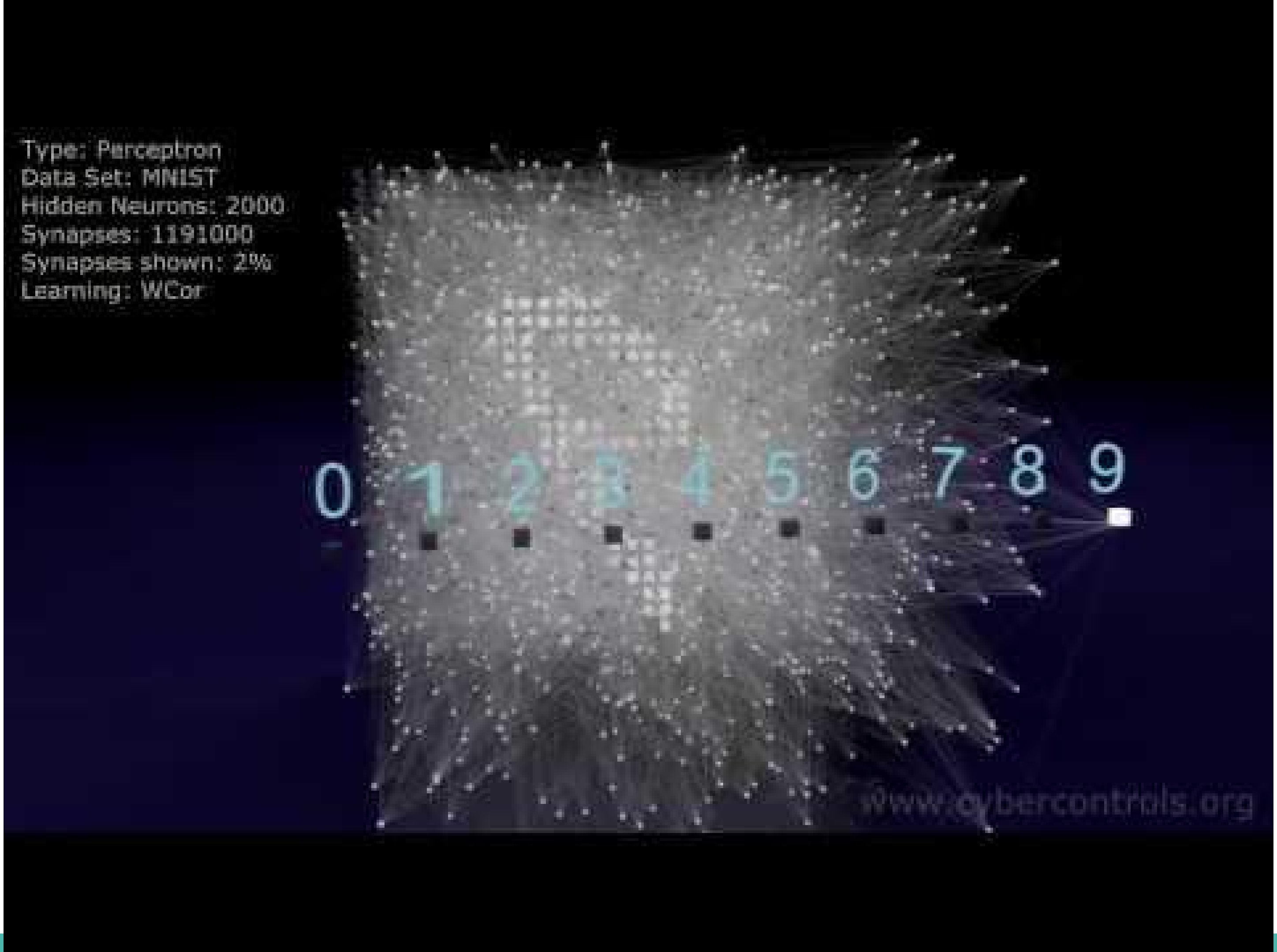
$$w_{new} = w_{old} - \eta \cdot \nabla_w J(w)$$

How do neural networks work? - learning





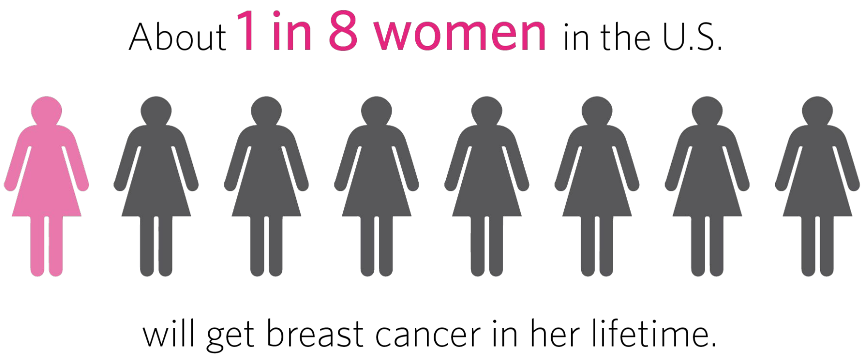

Theta1
Hidden layer
Theta2
Output layer
Y
Input layer 39
1
2
3
4
5
6
7
8
9
10
0.5 1 1.5
1
2
3
4
5
6
7
8
9
10
0.5 1 1.5


Type: Perceptron
Data Set: MNIST
Hidden Neurons: 2000
Synapses: 1191000
Synapses shown: 2%
Learning: WCor
0 1 2 3 4 5 6 7 8 9
www.cybercontrols.org

Deep learning for breast cancer detection

Overview

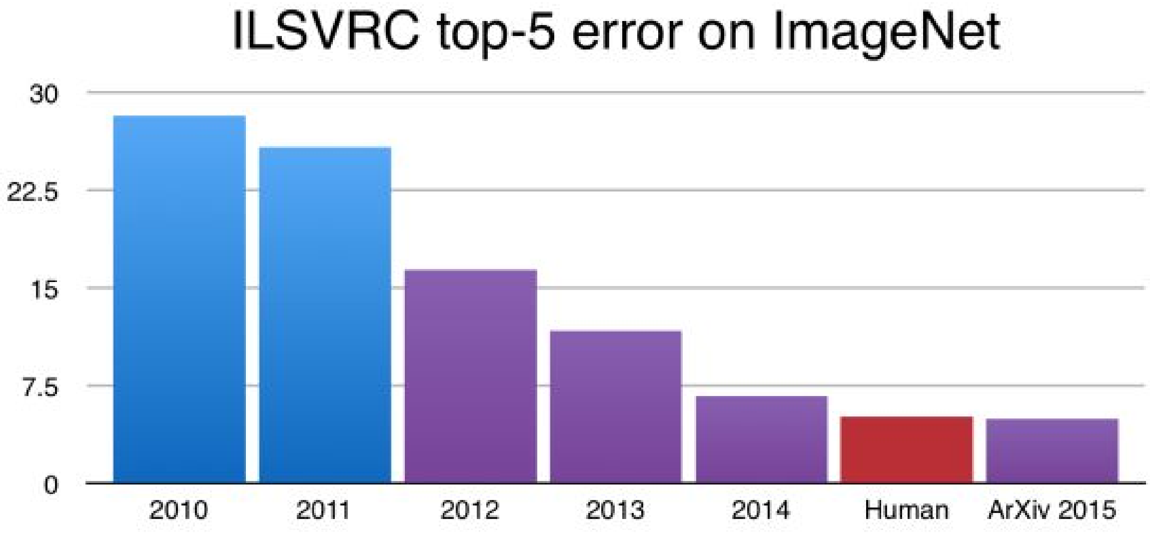
- Statistics
  - 266,120 cases annually in the US
  - 99% survival for local cancer
  - 26% survival for metastatic cancer
  - Early detection is crucial



- US recommends women receive annual mammograms starting at age 50
- **Problem:** radiologists are imperfect
  - False positives: Over a 10-year period, half of women receiving annual mammograms will be told they have cancer
  - False negatives: 1 in 5 cases of breast cancer are missed by radiologists
- **Goal:** Use deep learning to improve cancer detection in mammography

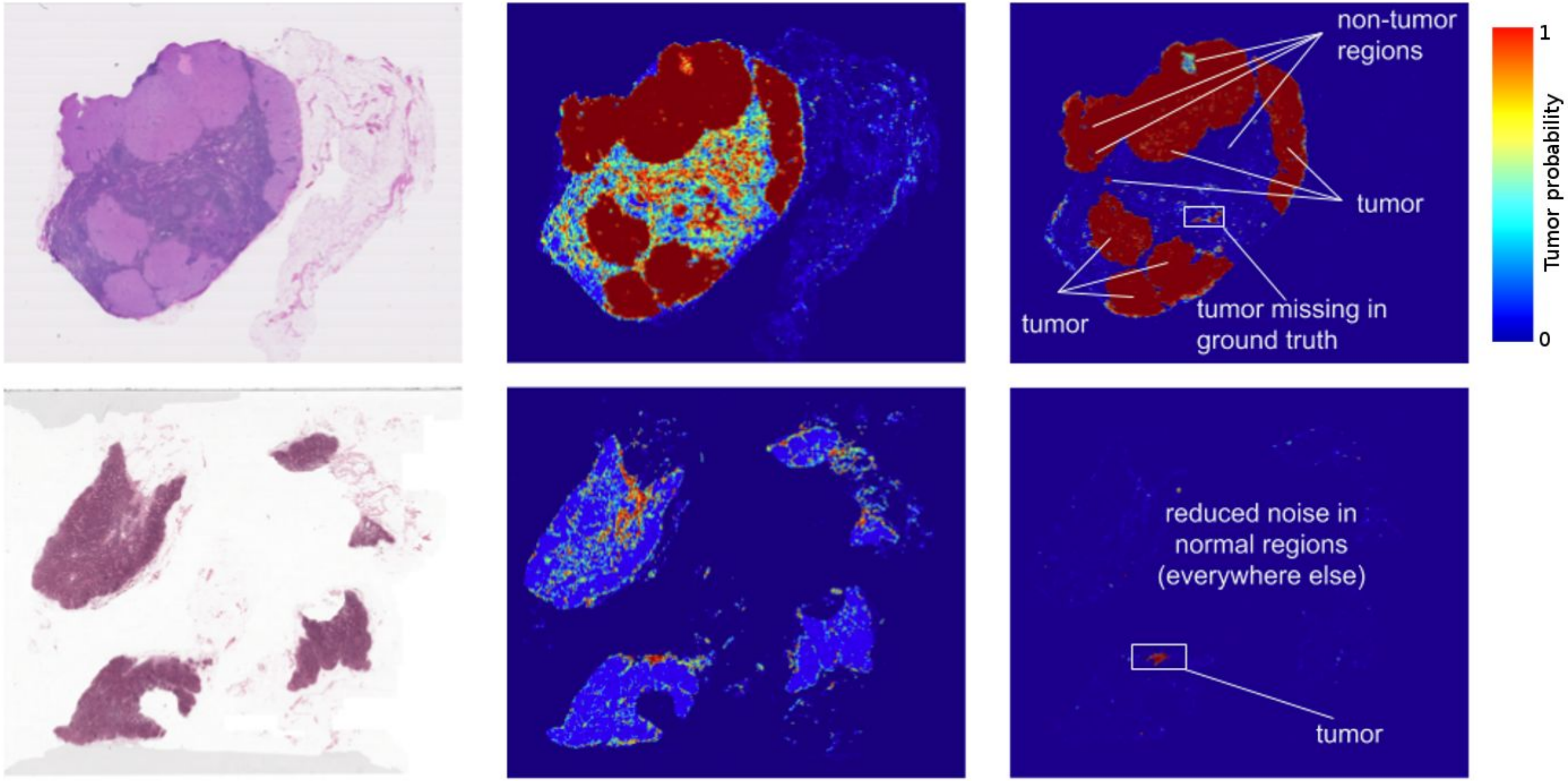
Related work - object detection

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
  - 1.2 million images, 1000 categories



Related work - object segmentation

Related work - cancer pathology



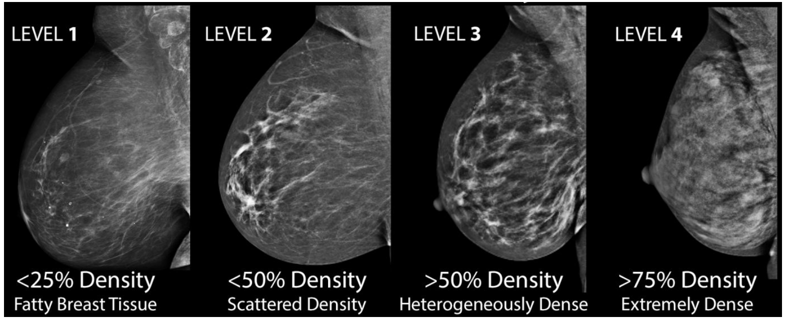
What makes mammography hard?

- Large images
  - 3000 x 4000 pixels rather than 256 x 256 pixels
  - GPU memory is limited
- Small region of interest
  - Cancers are typically 200 x 200 pixels
  - That’s only 0.3% of the image!
  - Non-cancer pixels significantly outweigh cancer pixels
- Masses can be benign or malignant
- One view is not enough
  - Radiologists see 2 or more views of the same breast
- Current mammogram is not enough
  - Radiologists rely on prior mammograms for comparison

How do we solve these problems?

- Start with an easier task
  - Predict risk of cancer
    - Global rather than local image feature
  - Use breast density as a proxy for risk
  - Standard methods achieve excellent results
- Proceed to the real task
  - Detect cancer itself
  - Current methods: Solve image size problems with patch-aggregator model
  - Future work: Use multiple views, prior images

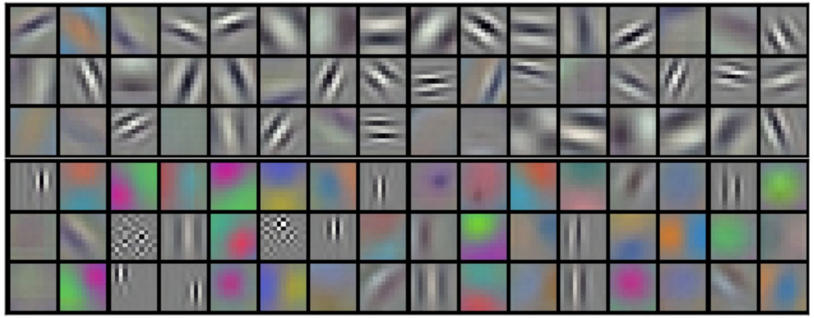
Density prediction - overview

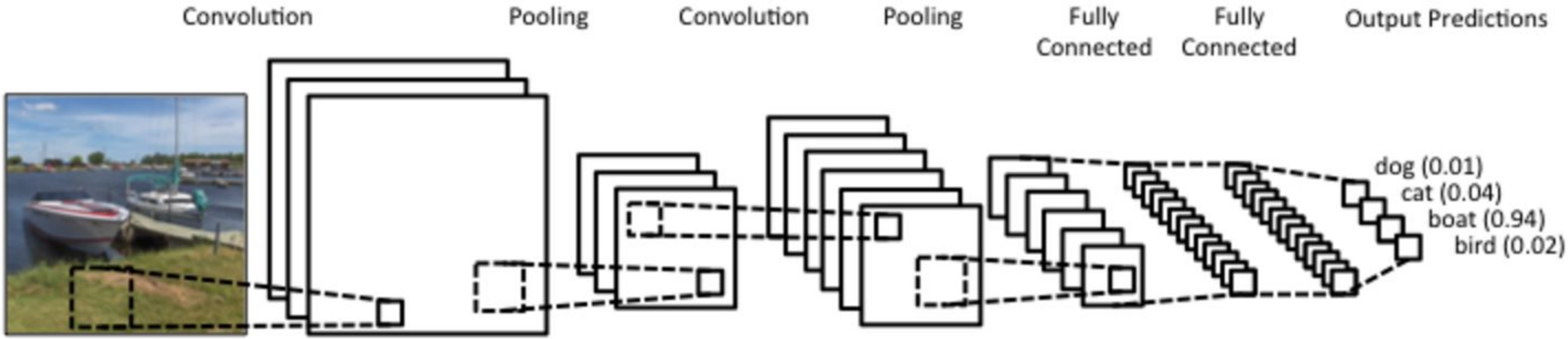


- Density as a risk factor
  - Women with high density breasts are 4-5 times more likely to get breast cancer than women with low density breasts
  - High density makes detecting cancer on mammograms more difficult
- Legislation: Many US states require that doctors notify women found to have dense breasts
- Density rating
  - 1-4 scale from low to high
  - 1,2 = low density
  - 3,4 = high density
- Highly subjective: Doctors across the US disagree on density ratings
- **Our aim:** Provide density ratings consistent with a top hospital (MGH)

Density prediction - methods

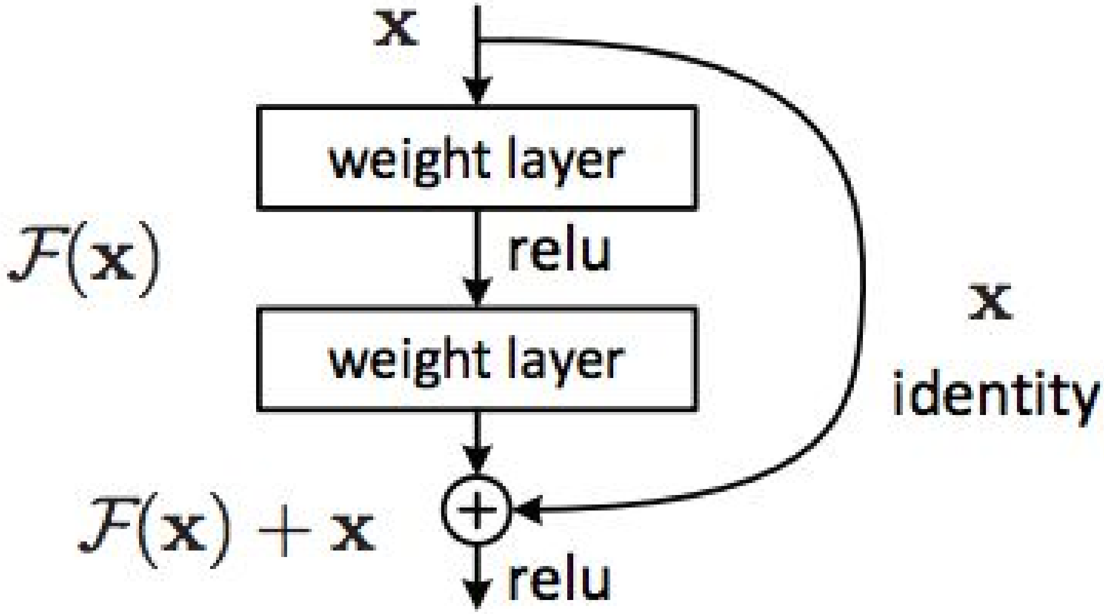
Convolutional Neural Network (CNN)

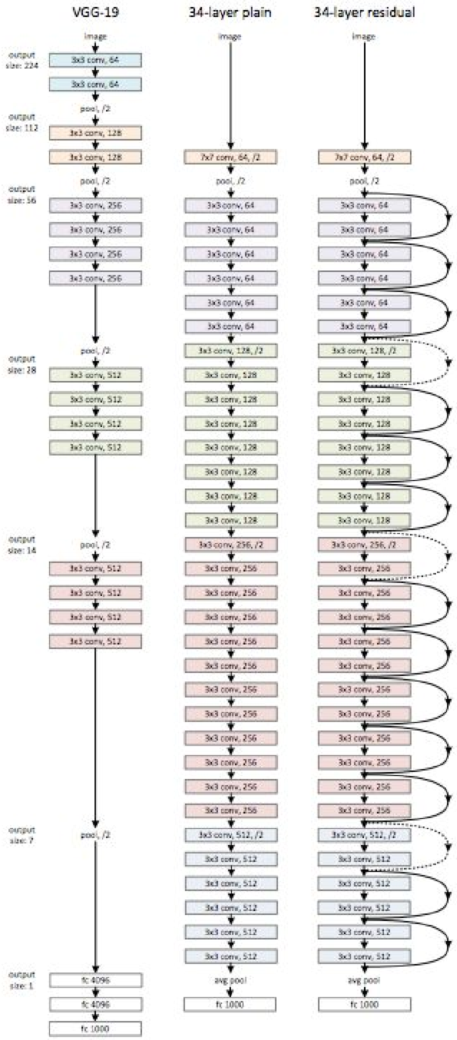




Density prediction - methods

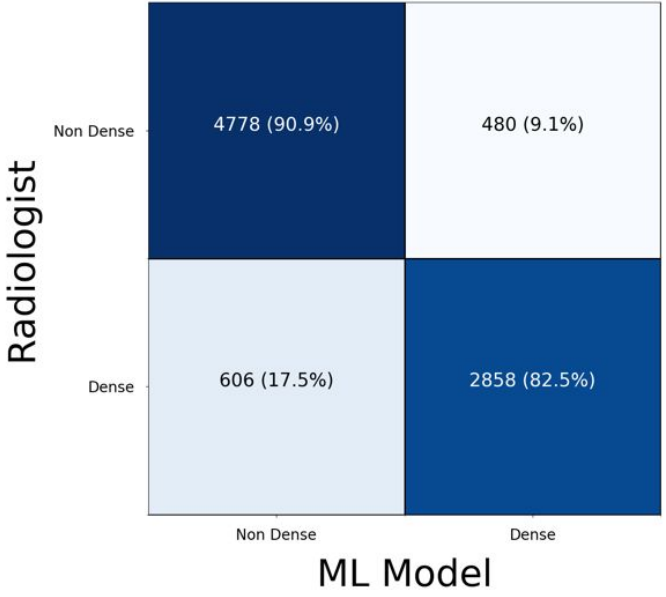
Deep Residual Network (ResNet)

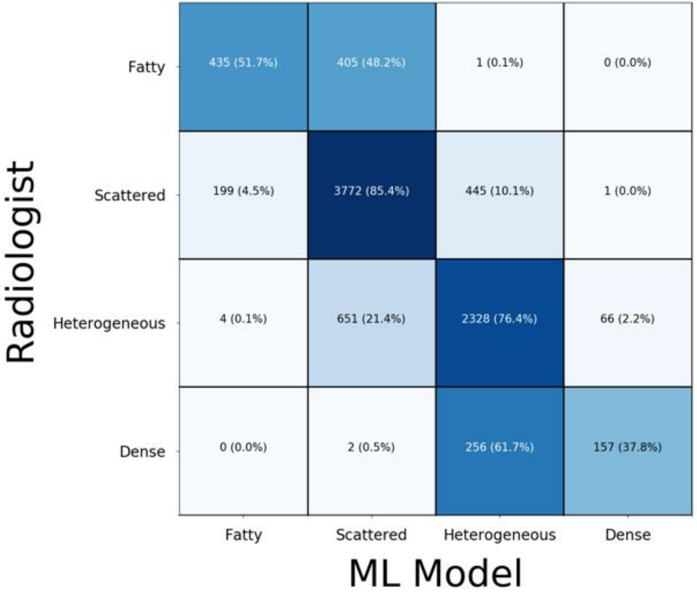




Density prediction - results (preliminary)

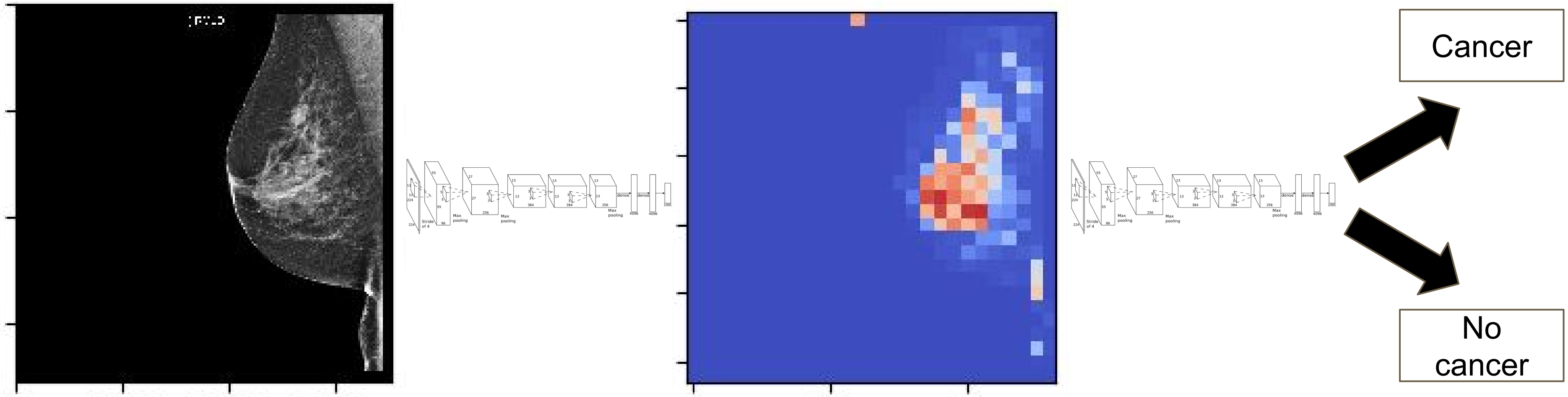
- Low vs. high: 85.0% model accuracy vs. 80.5% human accuracy
- 4-way: 71.0% model accuracy vs. 66.3% human accuracy





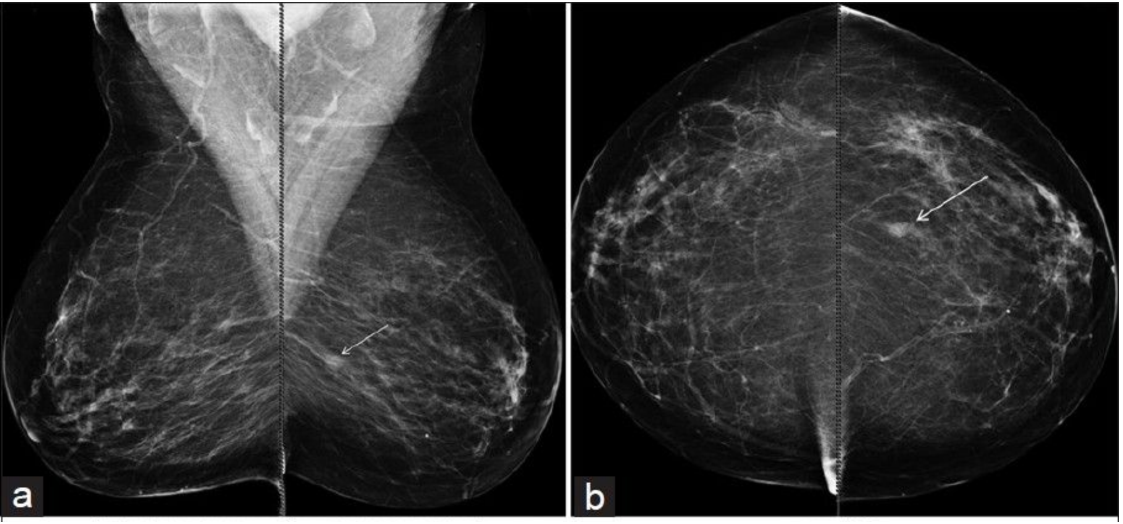
Cancer detection - methods

Patch-aggregator model

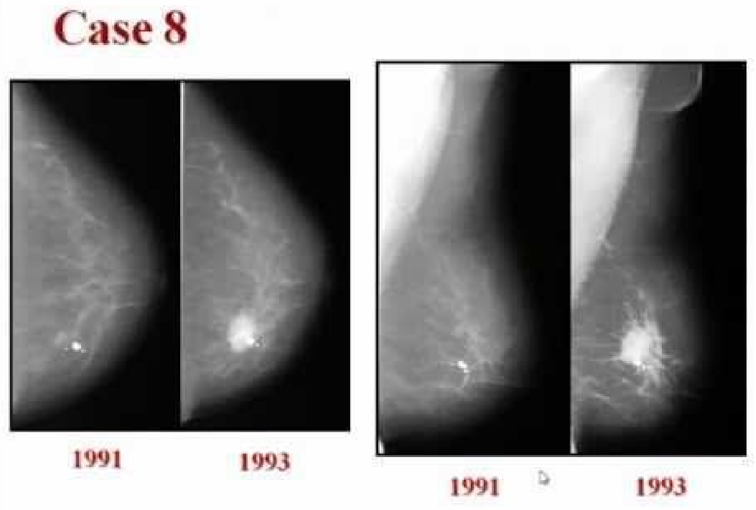


Cancer detection - future work

Multi-view



Multi-time



Summary

- Machine learning recap
- What is deep learning?
- How do neural networks work?
- Deep learning applications
- Case study
    - Deep learning for breast cancer detection
- Questions?
- Email: swansonk@mit.edu