Deep Learning Tutorial

Aug. 2016

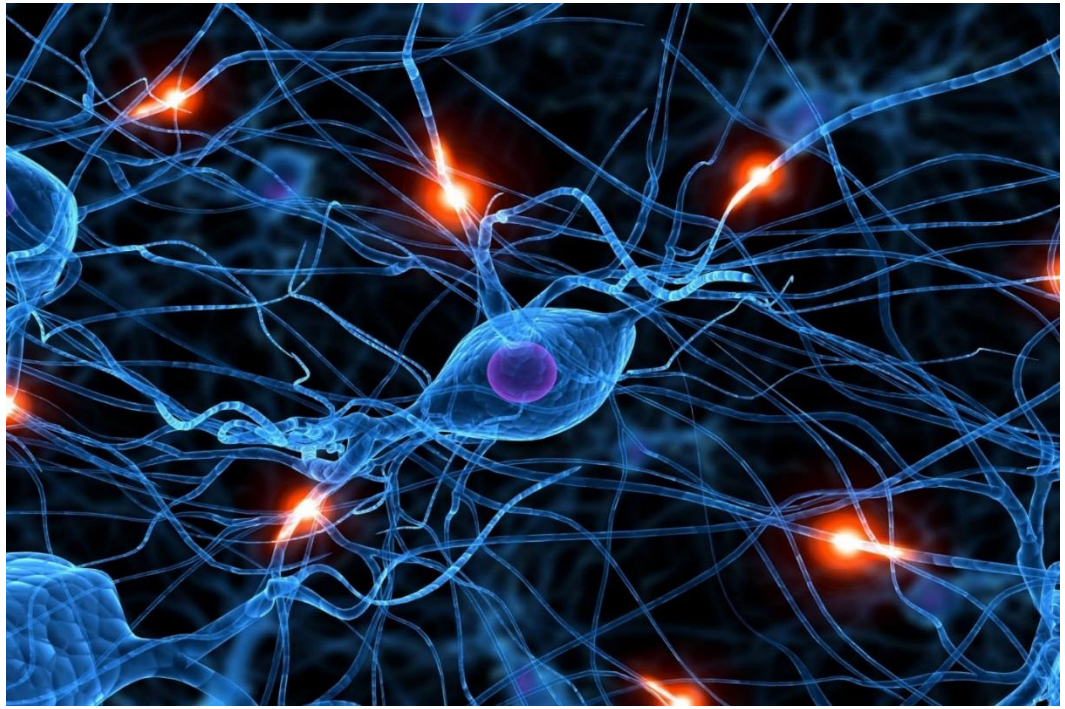
Machine Learning

- The science of getting computers to act without being explicitly programmed.

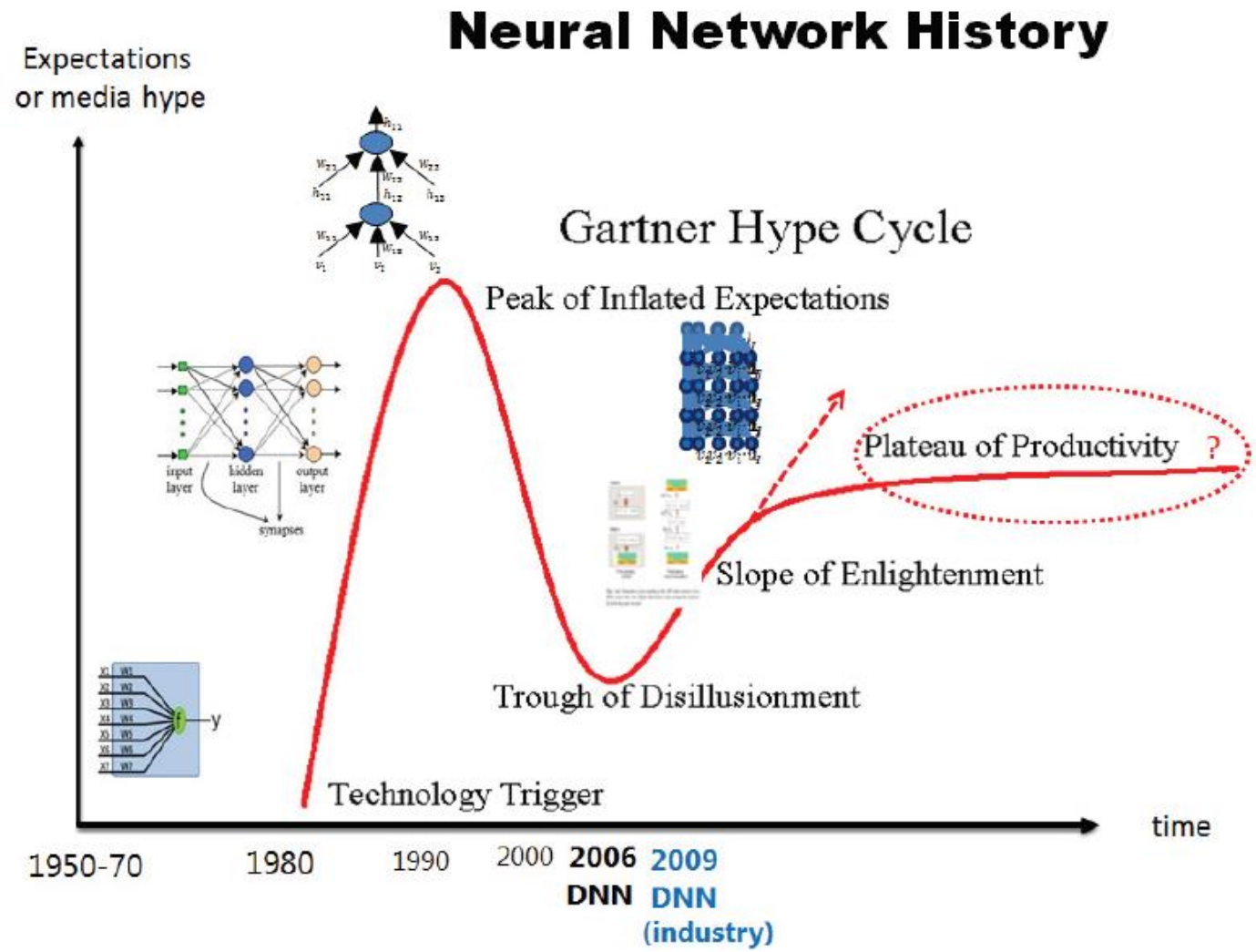
```
PROGRAM BiggerOfThree:
    Read A;
    Read B;
    Read C;
    IF (A>B)
            THEN IF (A>C)
                            THEN Print A;
                            ELSE Print C;
                        END IF;
            ELSE IF (B>C)
                            THEN Print B;
                            ELSE Print C;
                    END IF;
    END IF;
END.
```

Neural Networks

- Inspired from human brain!!! (really ;))



History


Neural Network History
Expectations or media hype
Gartner Hype Cycle
Peak of Inflated Expectations
Plateau of Productivity ?
Slope of Enlightenment
Trough of Disillusionment
Technology Trigger
time
1950-70
1980
1990
2000
2006 DNN
2009 DNN (industry)
input layer
hidden layer
output layer
synapses

Applications

- Image recognition

Image/Video

- Speech recognition

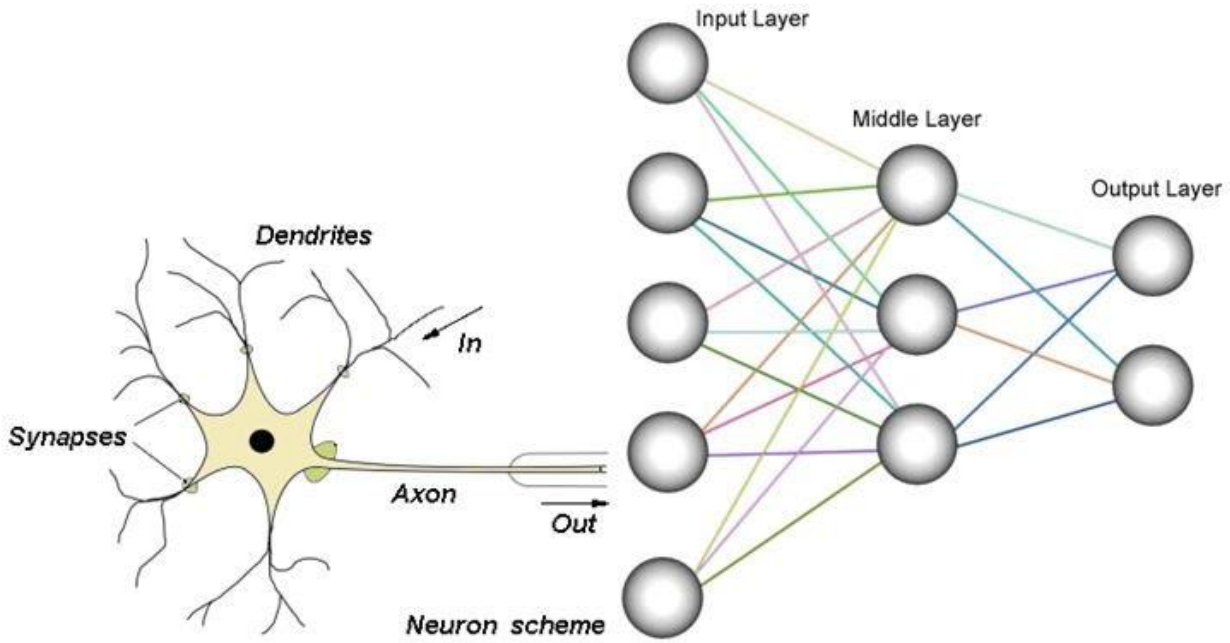
Audio

- Web search

Text

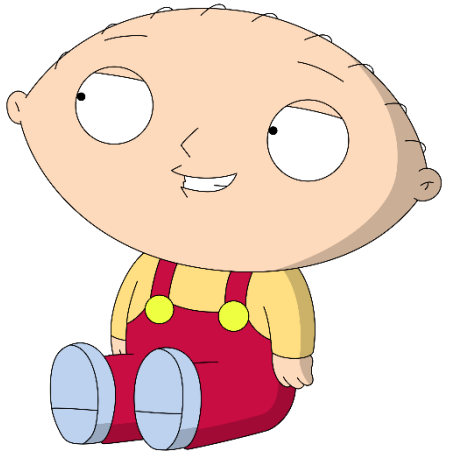
Deep Learning

- A set of algorithms based on neural networks
  - **Convolutional neural networks**
  - Deep belief networks and auto-encoders
  - Recurrent neural networks



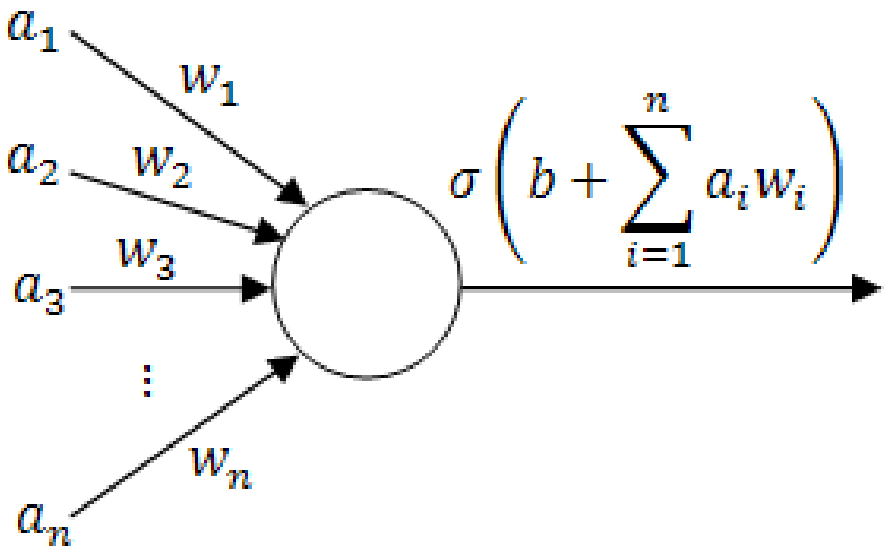
AI Paradigms

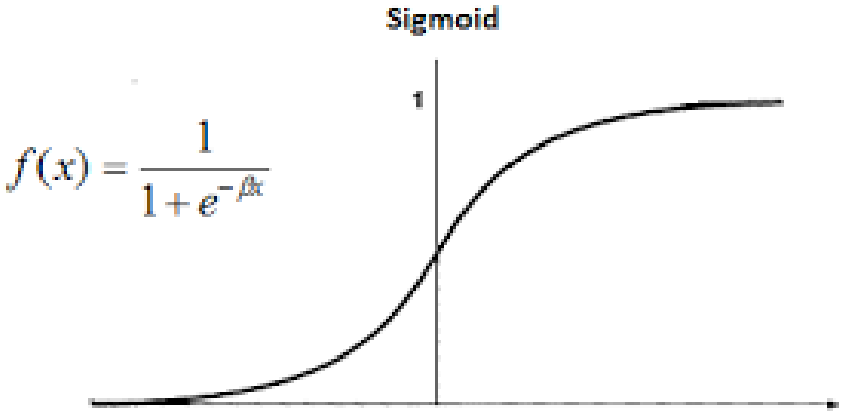
Supervised	Unsupervised
Require lots of examples with labels Most of current AI	Less examples, no/less labels Future AI



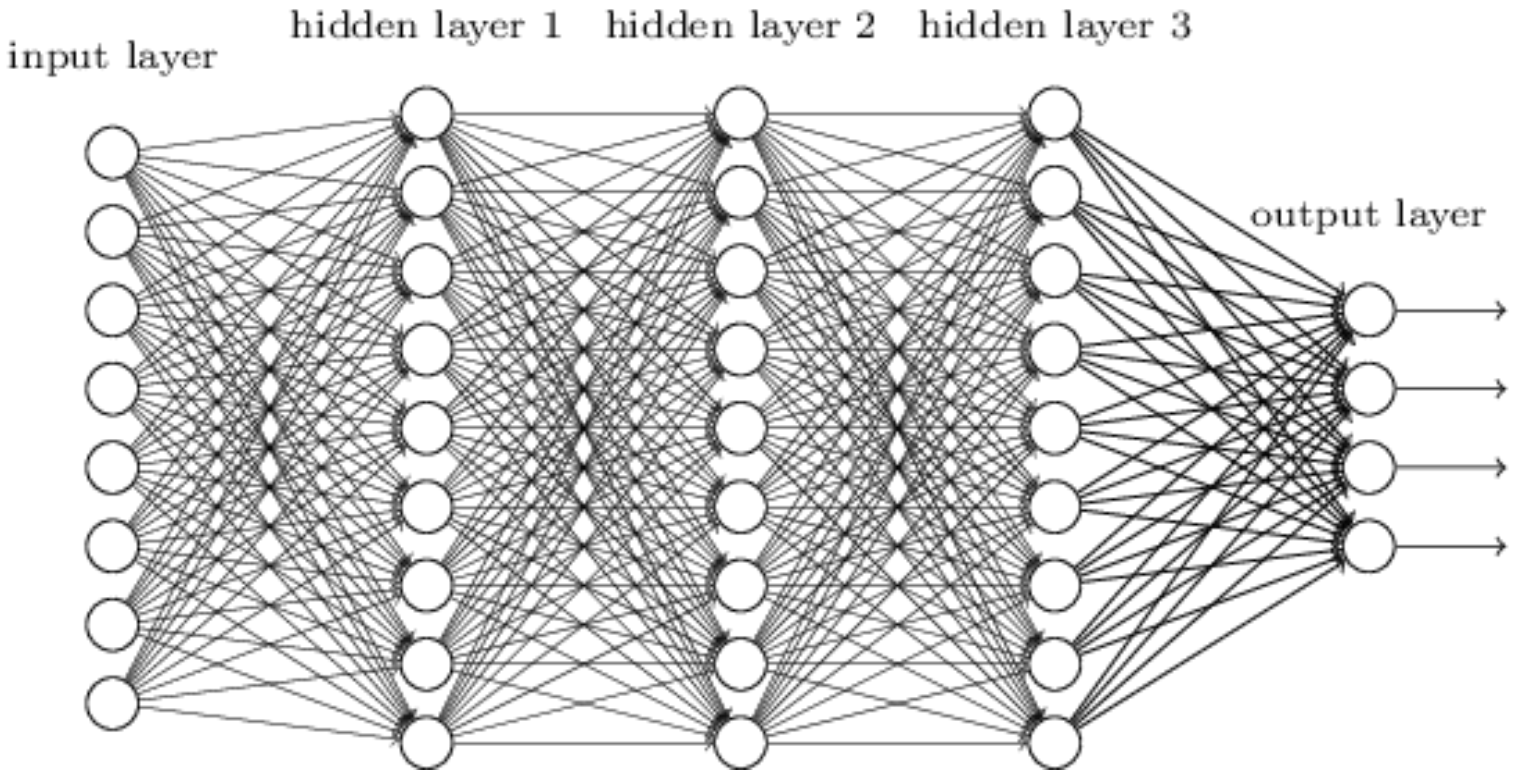
Perceptron

- Activation function
  - Sigmoid, ReLU, etc.

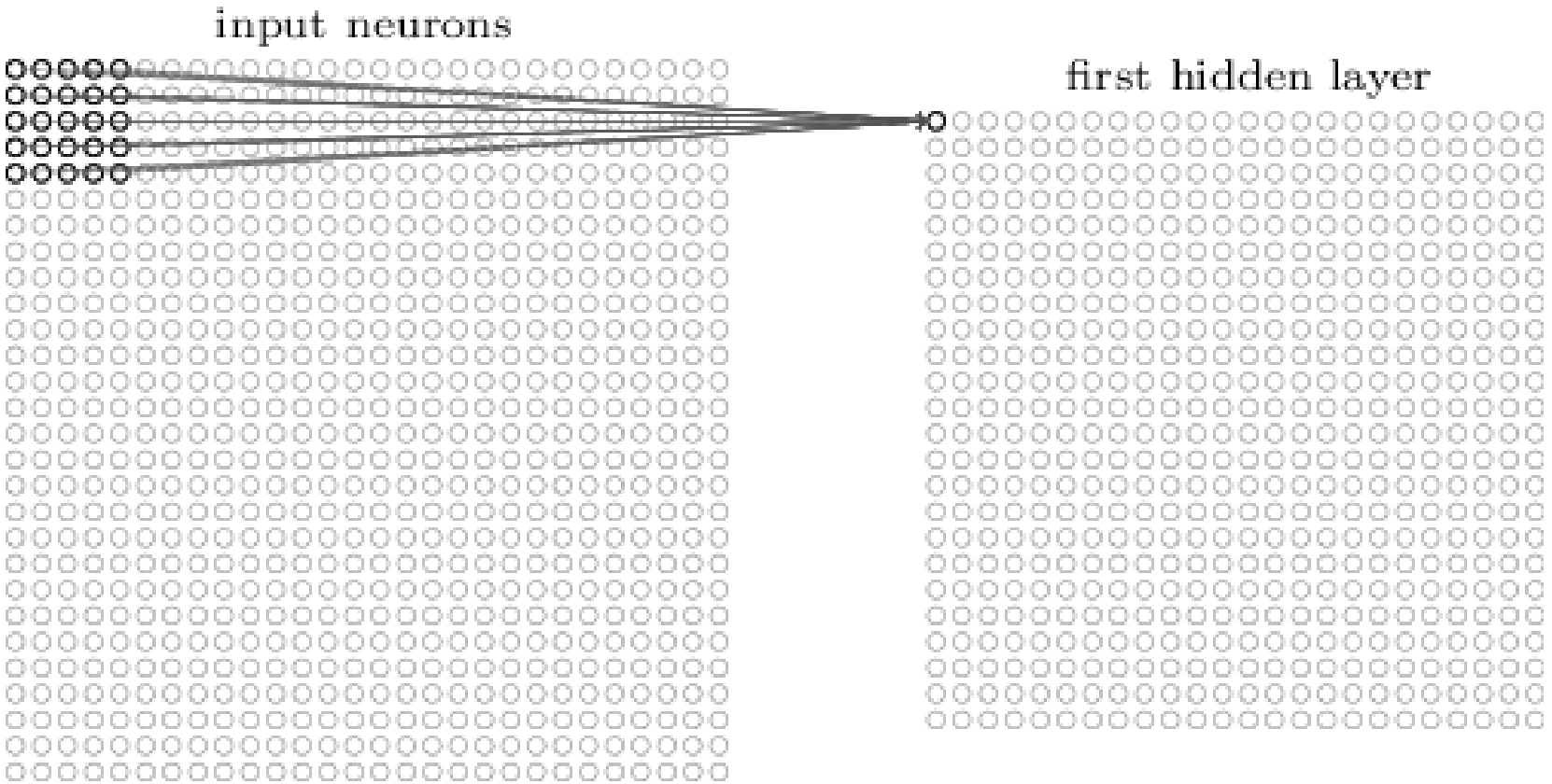




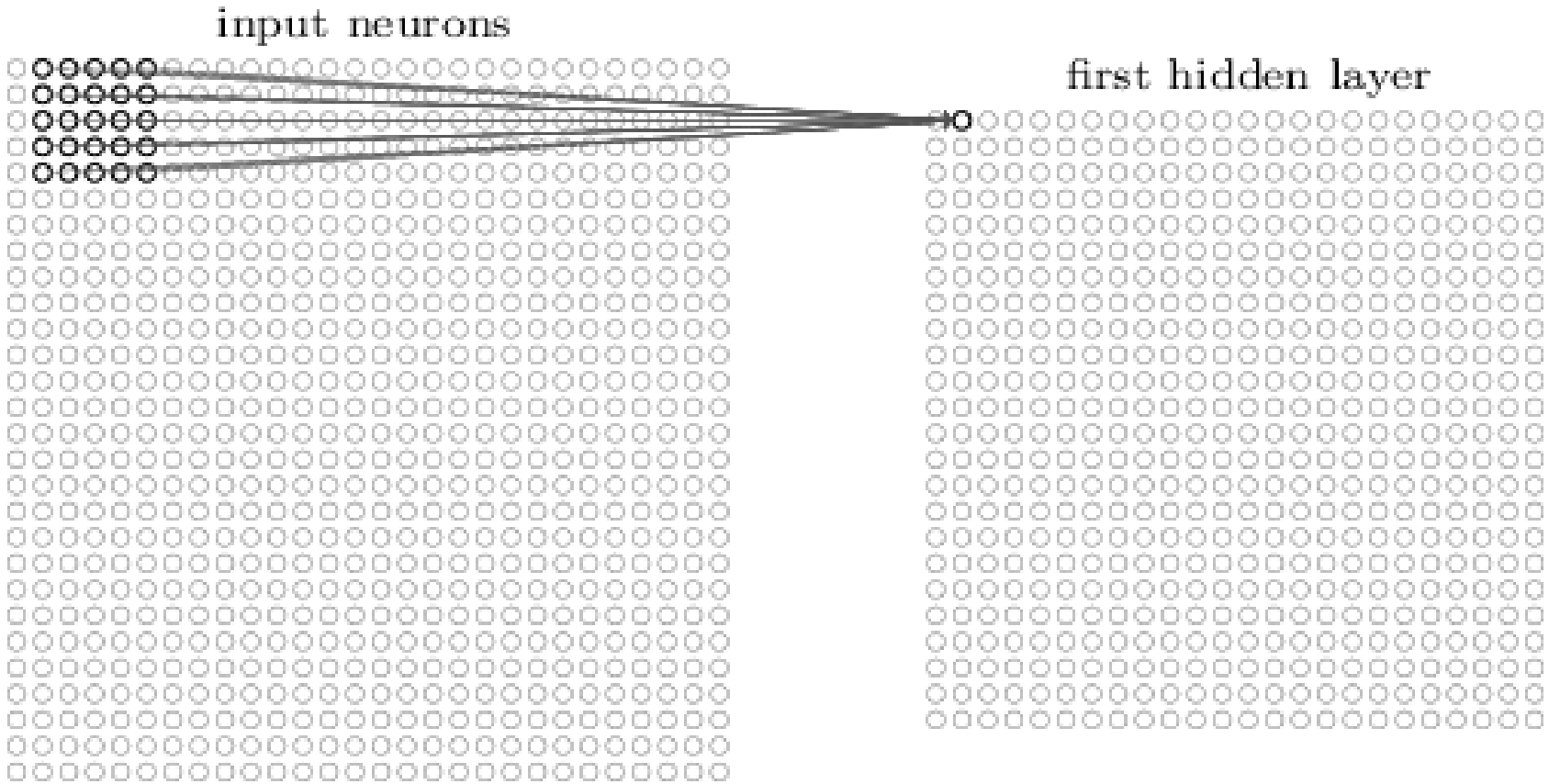
Deep Belief Networks



Convolutional Neural Networks (CNN)



Convolutional Neural Networks (CNN)



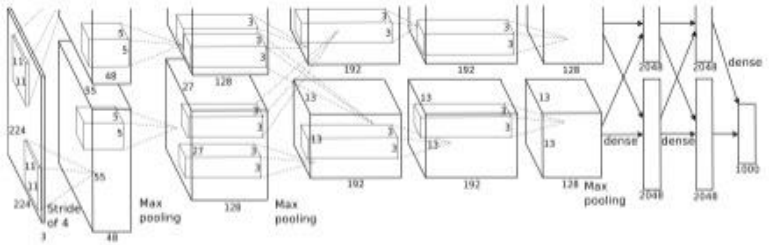
Convolutional Neural Networks

$$\sigma\left(b + \sum_{l=0}^{4}\sum_{m=0}^{4} w_{l,m} a_{j+l,k+m}\right)$$

Training

- **Backpropagation**

Forward: inference $f_W(x)$



"espresso" + loss

$\nabla f_W(x)$ Backward: learning

Loss Functions

- Depending on application
  - Mean squared error (MSE)
  - Binary cross entropy
  - Multi-class class entropy

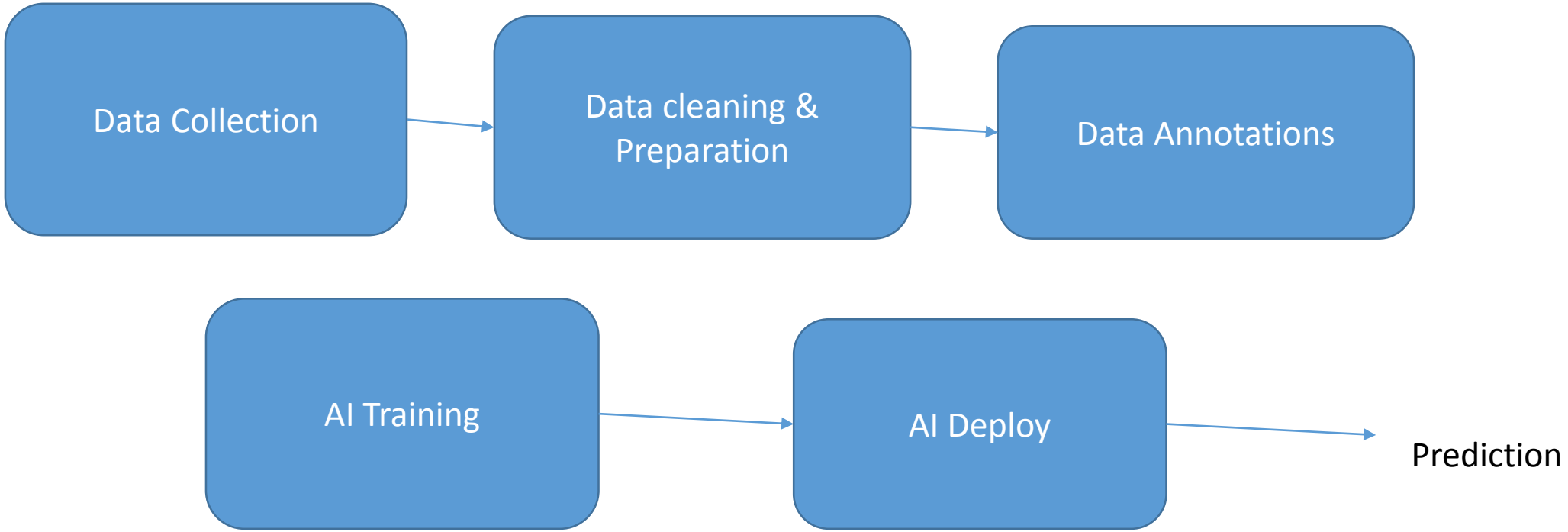
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

$$Entropy(X) = -p_{Yes}\log_2 p_{Yes} - p_{No}\log_2 p_{No}$$

Optimization Method

- **Stochastic Gradient Descent (SGD)**
- Adam
- AdaDelta
- Nestrov

Deep Learning Pipeline

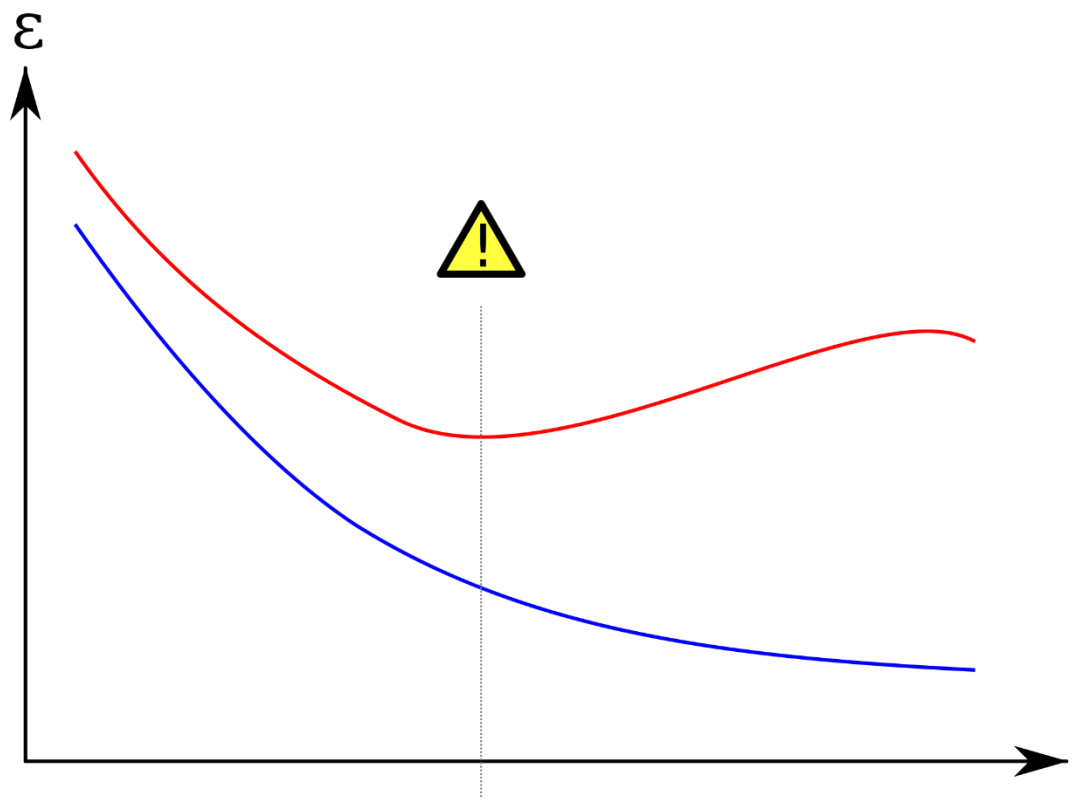


Deep Learning Frameworks

- Theano: Montreal
- Torch: NYC
- Tensorflow: Google
- **Caffe**: Berkeley
- **Keras** (Tensorflow,Theano)
- etc

Model Design

- Do not start from scratch
  - Use pre-designed models
- Start from small models
  - Filters, hidden layers, etc.
- Reduce complexity
  - Down-sample input/output
- Clean data
  - Remove outliers
- Cross validation
  - Train validation data split
  - Monitor training progress

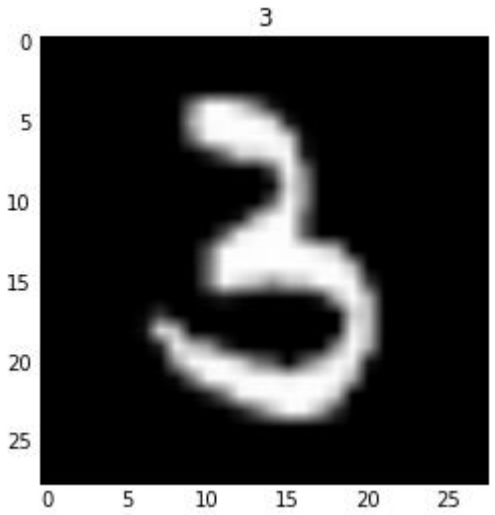


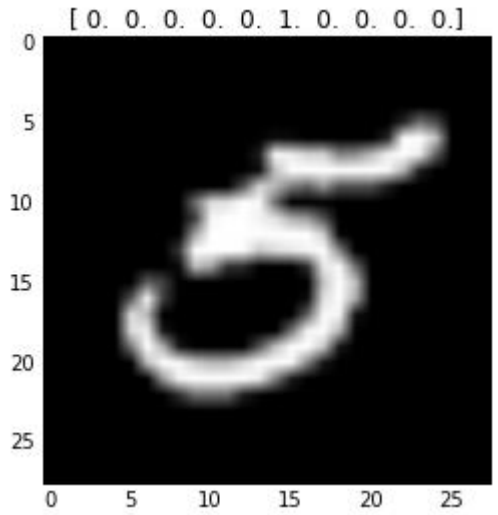
Keras

- Deep learning framework
  - Tensorflow
  - Theano

Example

- Digit recognition, MNIST





Digital Recognition

- Model