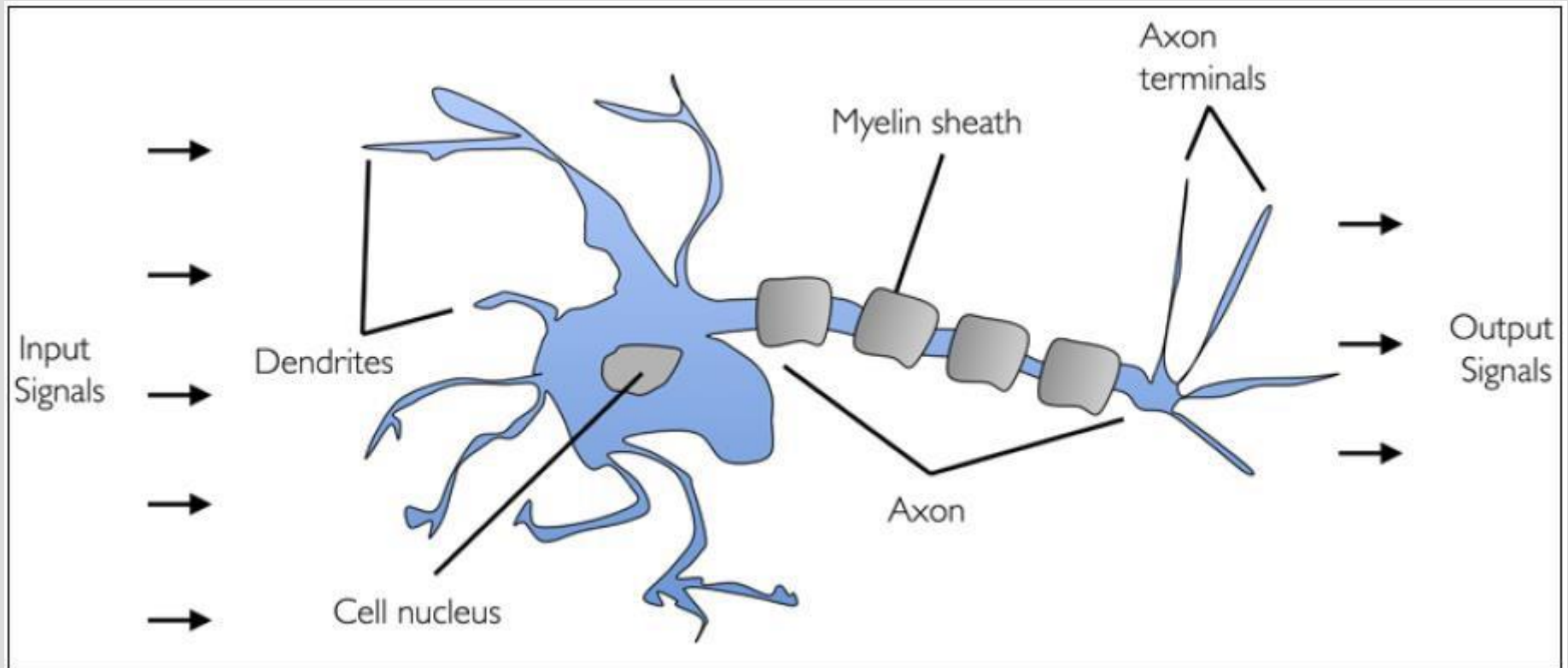


# Neural Networks

# Topics

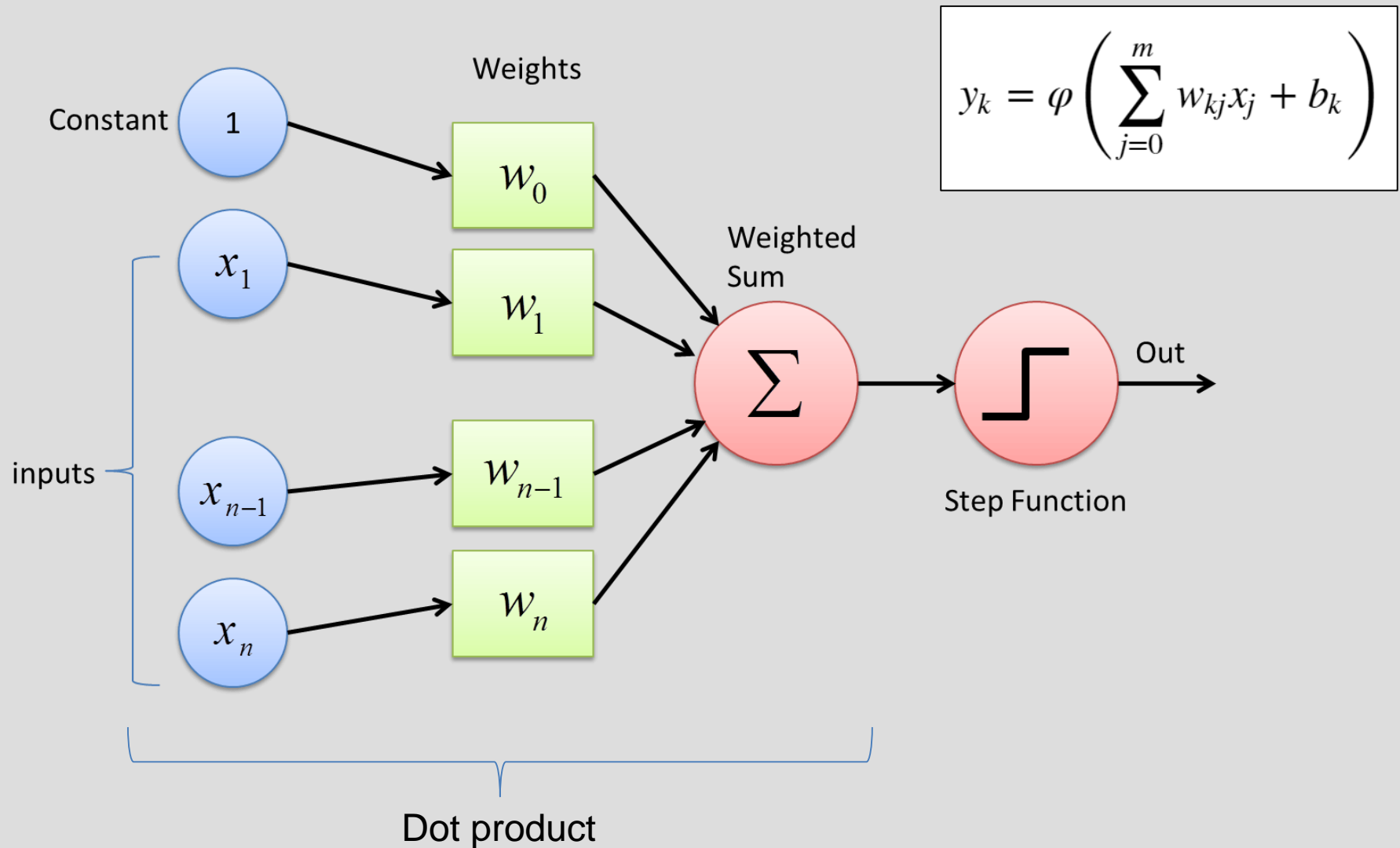
- Neuron
- Perceptron
- Neural Networks
  - Feed Forward
  - Back Propagation

# Biological Motivation

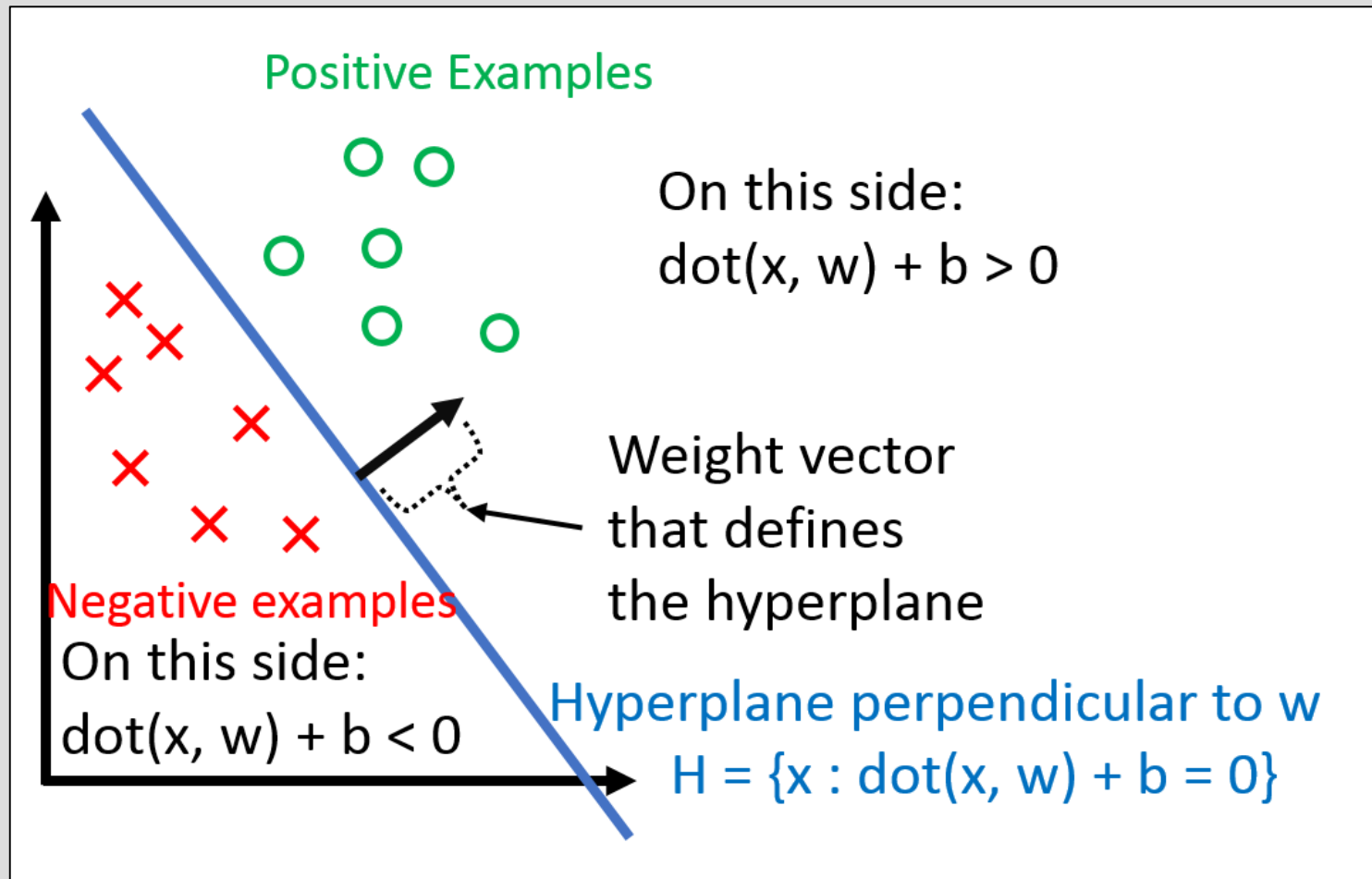


<https://www.simplilearn.com/what-is-perceptron-tutorial>

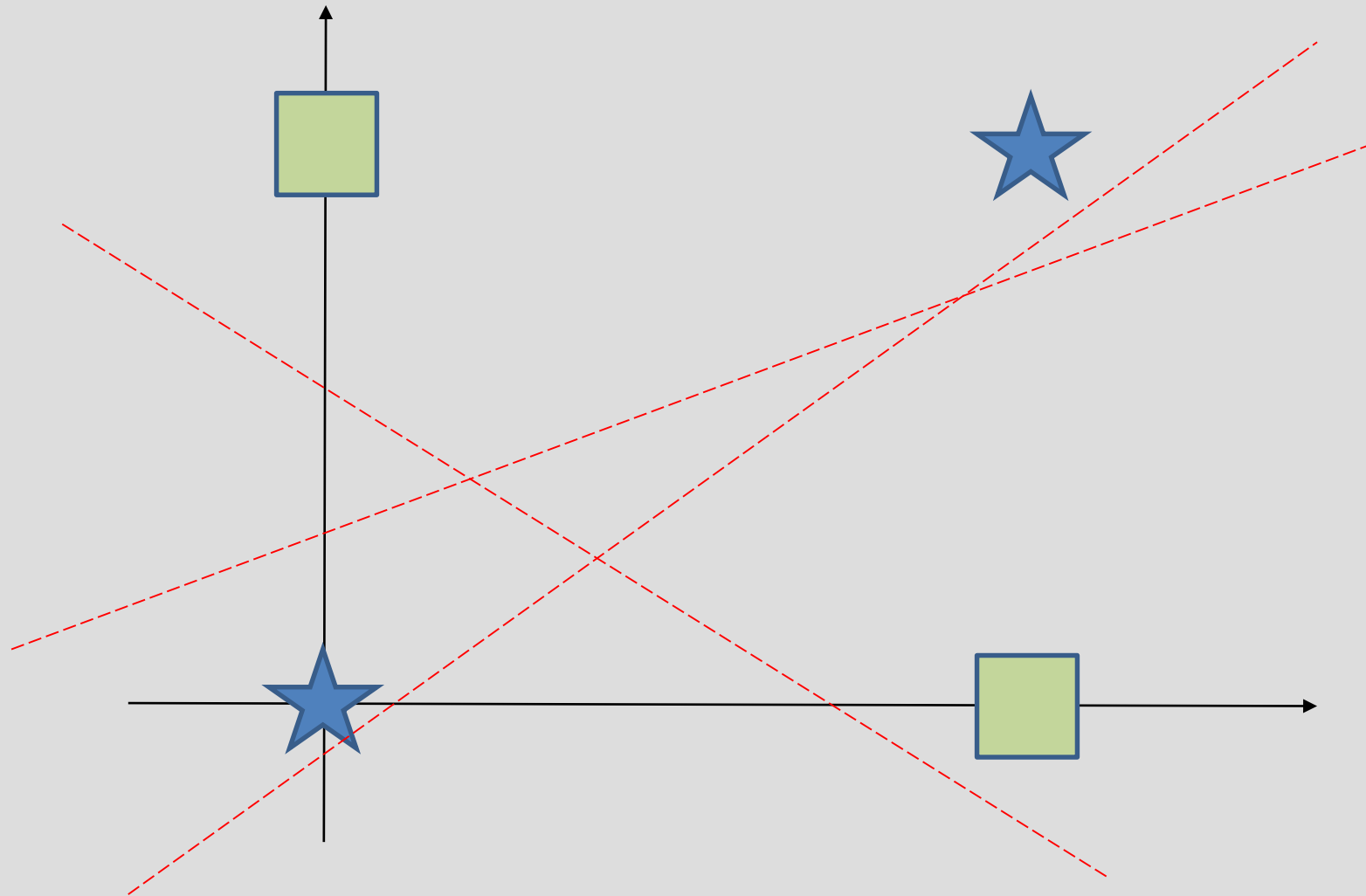
# Perceptron



# Geometric Interpretation



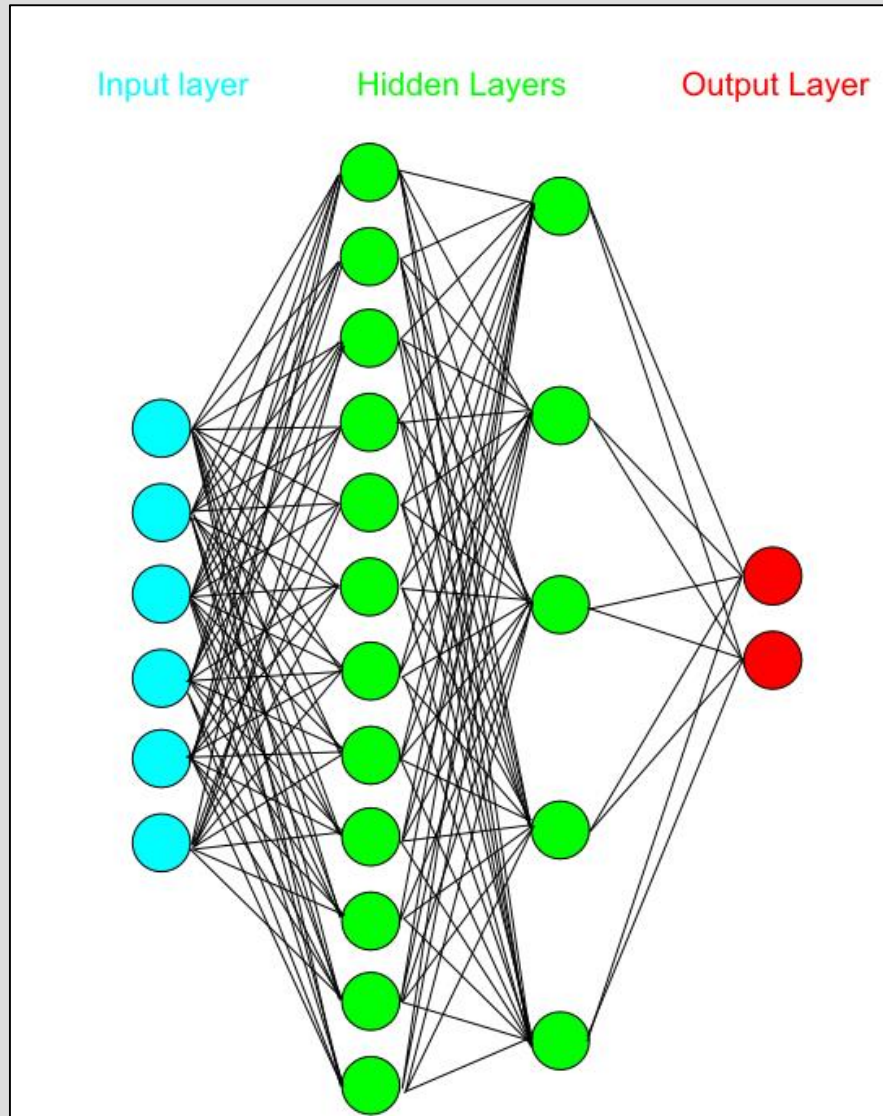
# XOR





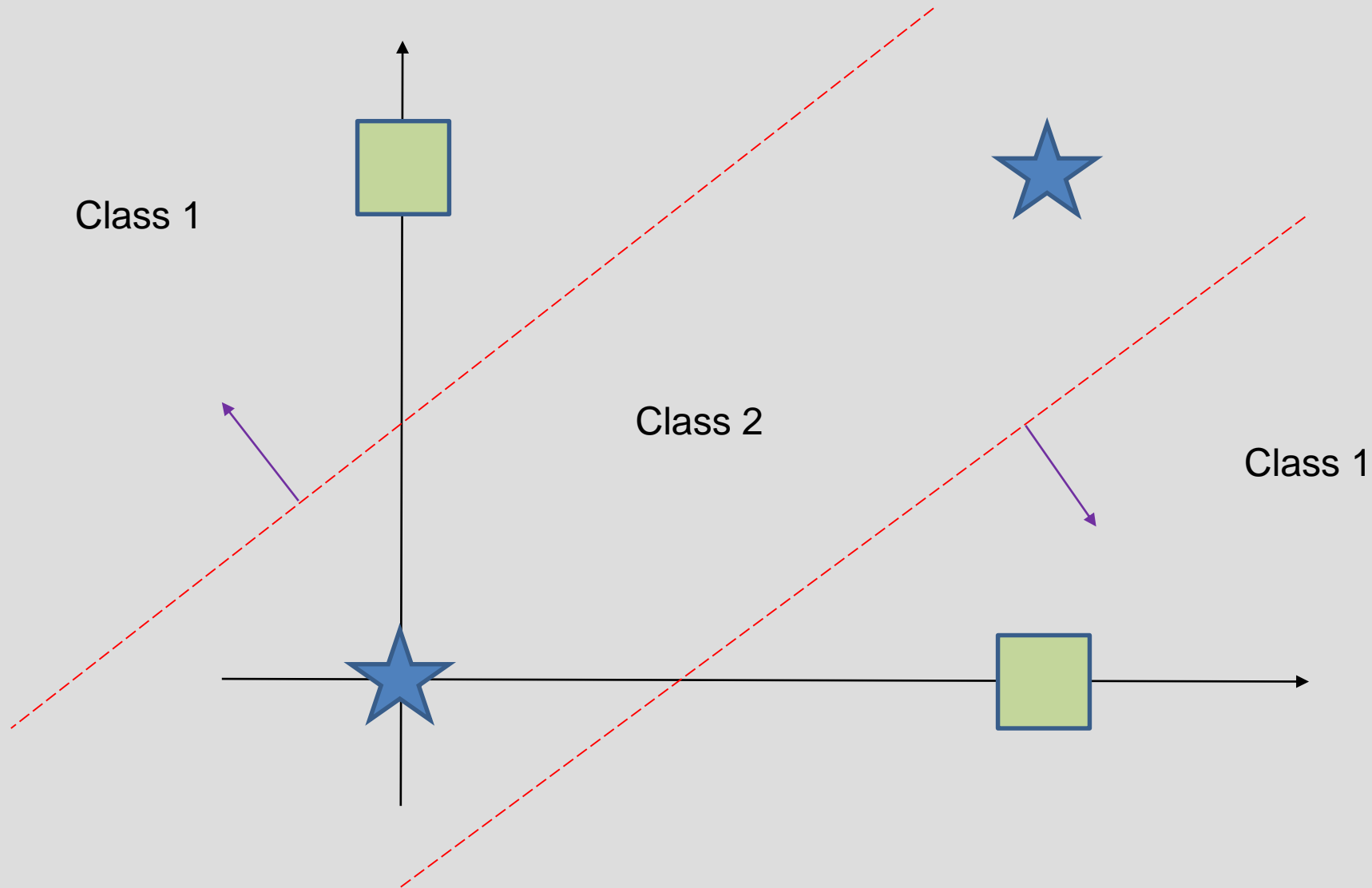
DATA SCIENCE  
& ANALYTICS

# Multi-Layer Perceptron



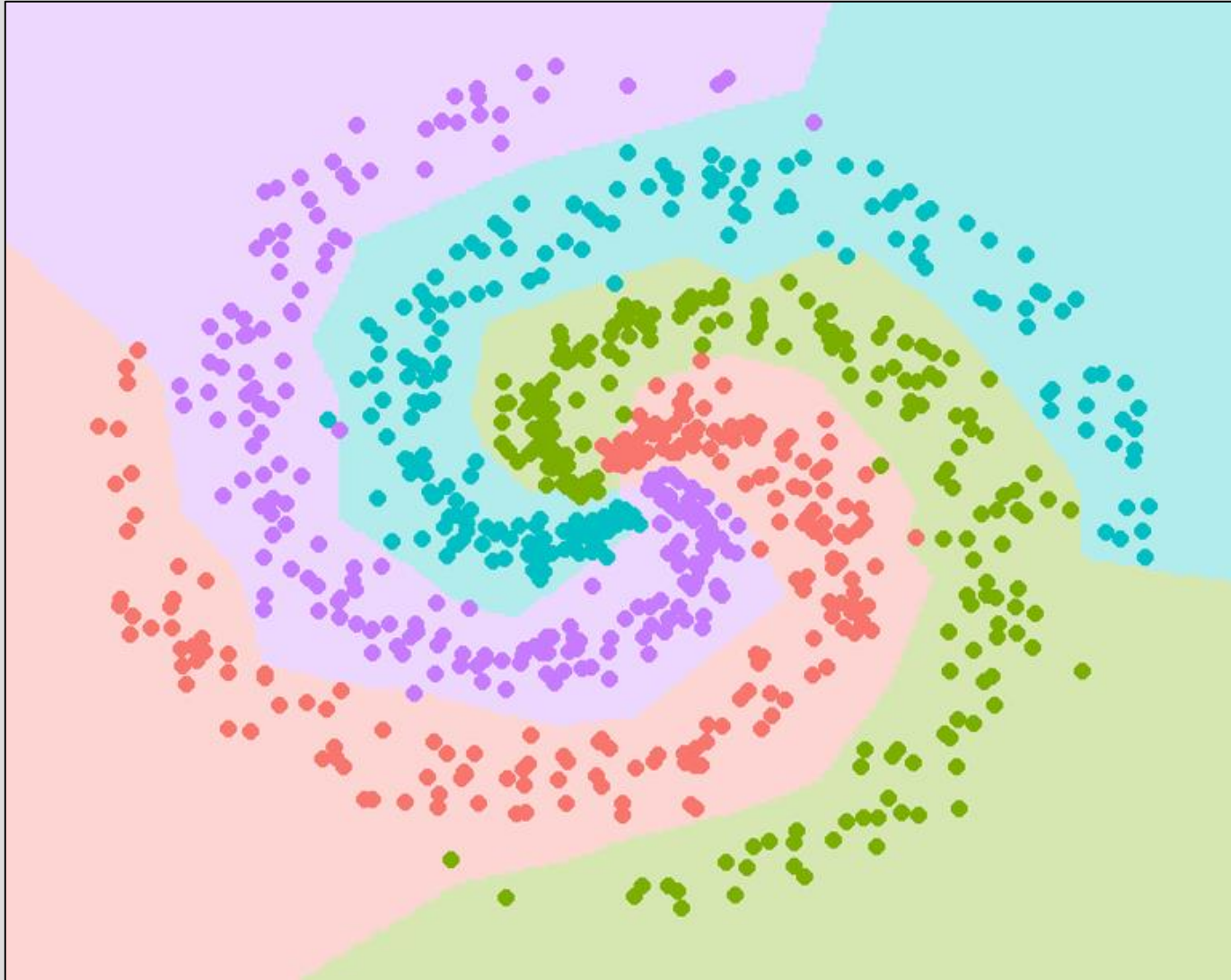
<http://blog.refu.co/?p=931>

# MLP and XOR



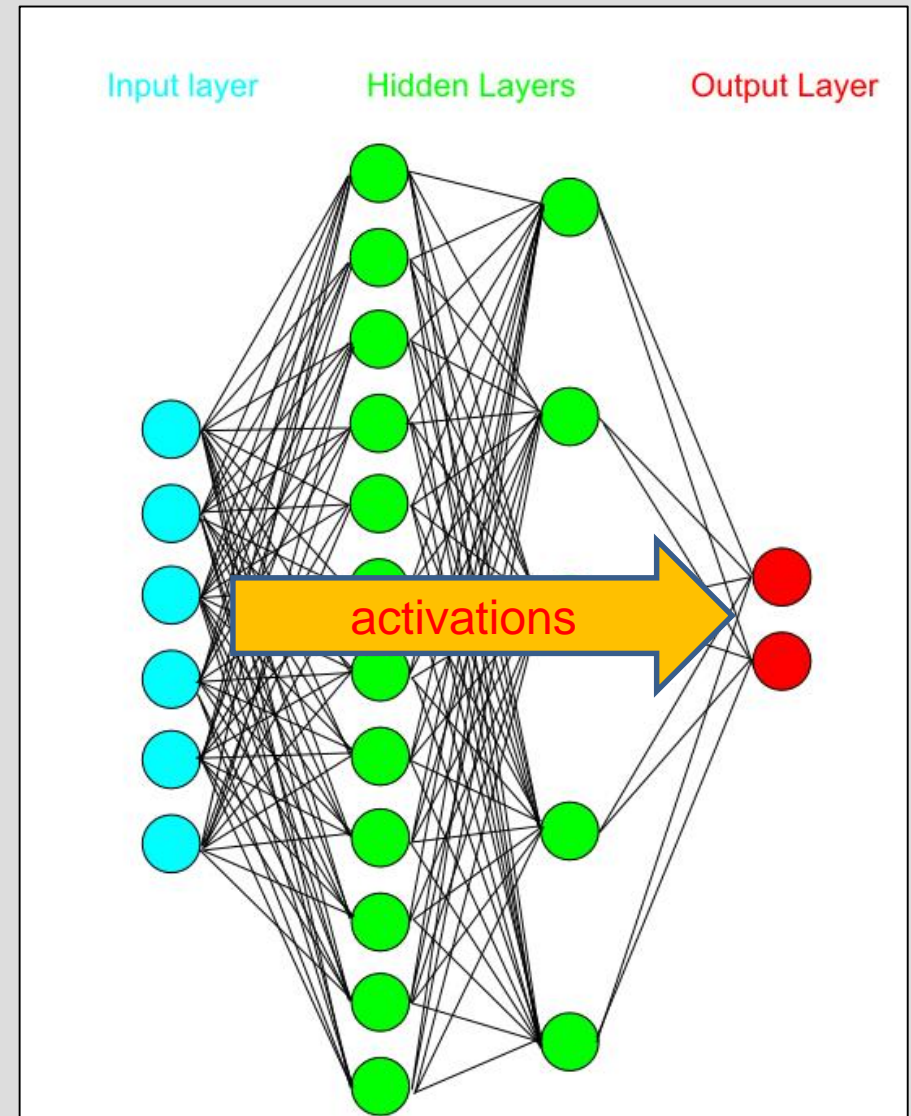


# (Overly!?) Complex Solutions



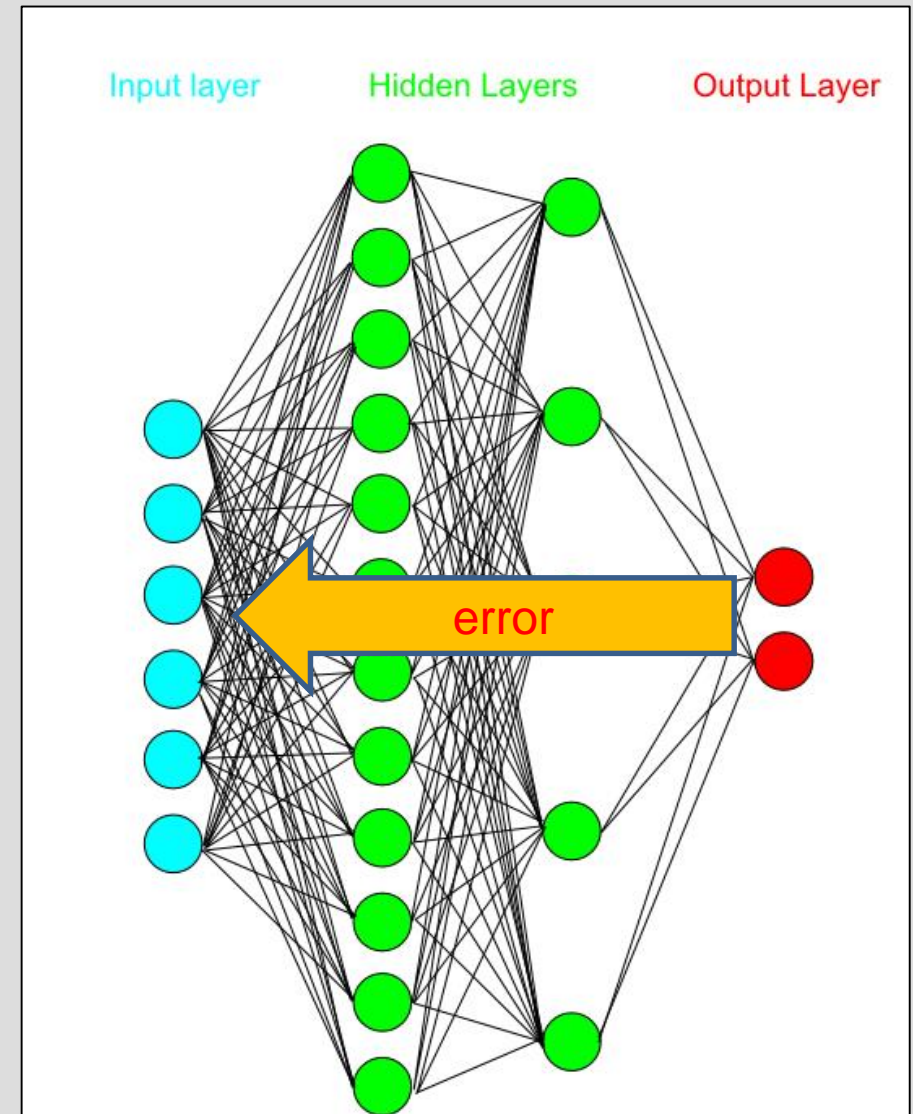
# Feed Forward

- Data comes in
- Calc perceptron
- Input to the next layer
- Calc perceptron
- ...
- Get “result”
  - Class 1 and 2
  - Do something!

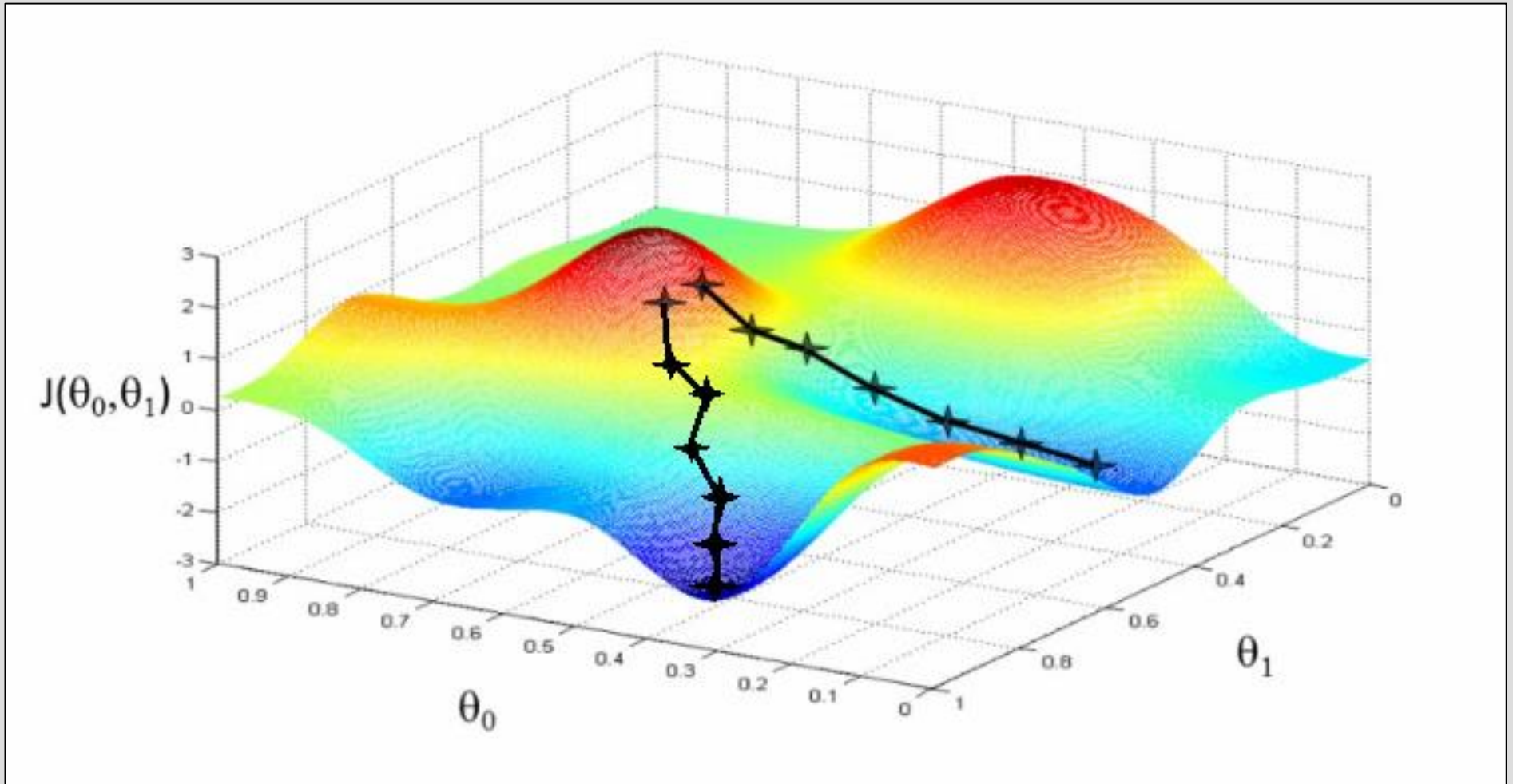


# Backpropagation

- Calc partial derivatives
- Calc error on data
  - SSE or MSE or ...
- Push back a layer
- Update parameters
- Push back a layer
- ...
- Done!

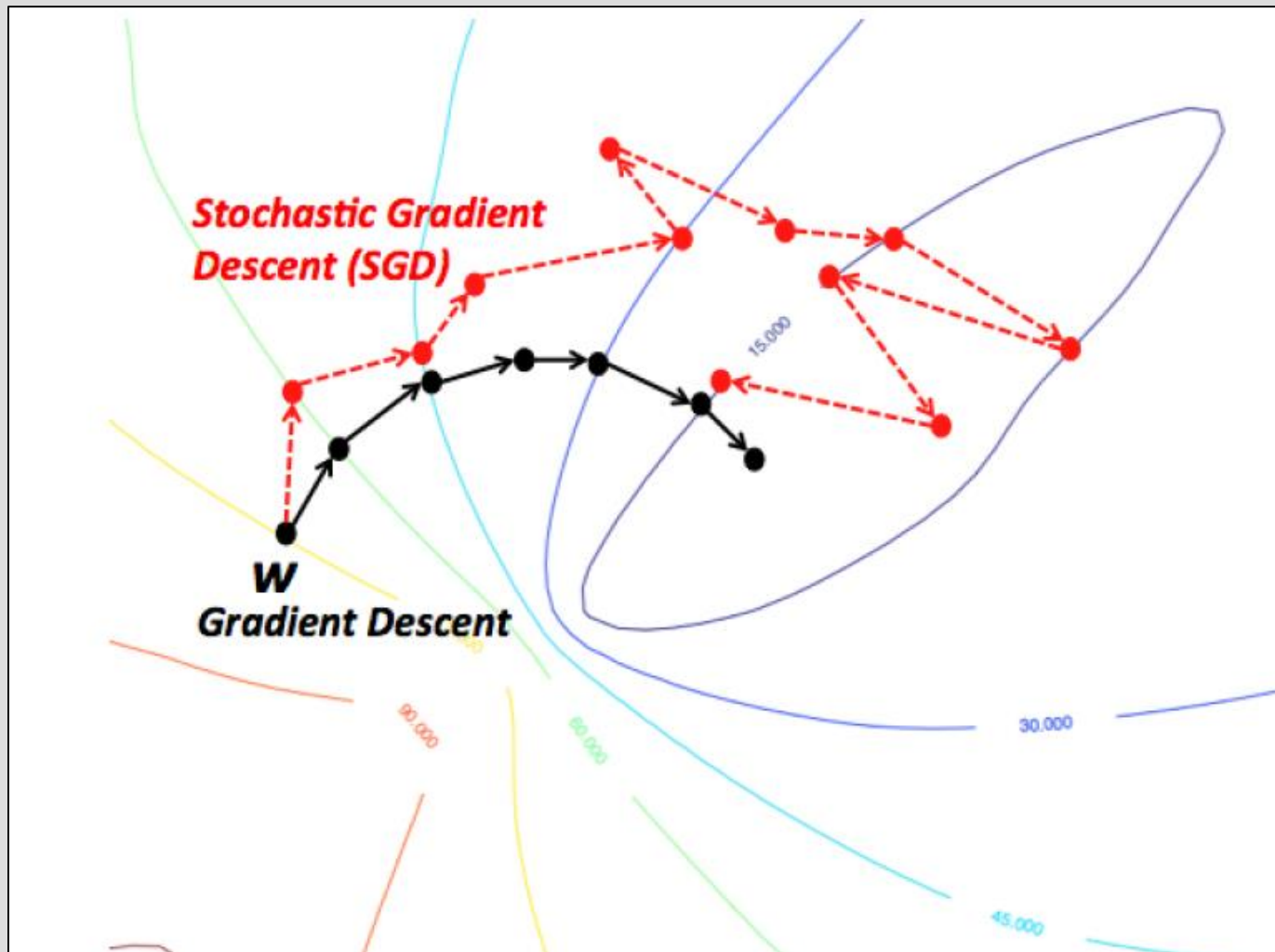


# Global and Local Minima



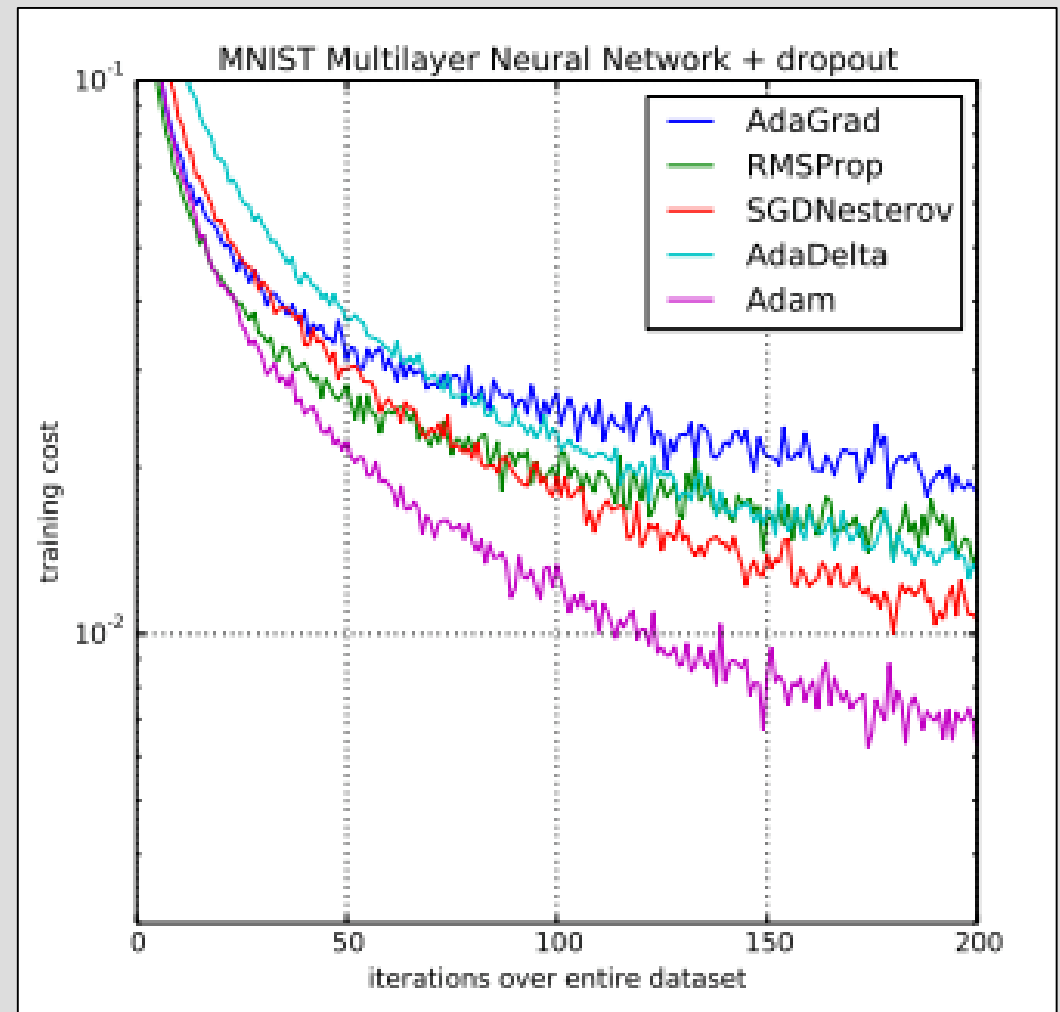


# Stochastic Gradient Descent



# Adam

- Adam is an adaptive learning rate method
- Combination of RMSprop and Stochastic Gradient Descent with momentum
- Many current DL packages use Adam for optimization



# Gradient Descent

- Batch
  - Computes gradient of cost function w.r.t. to the parameters  $\theta$  for entire training dataset

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta)$$

- Stochastic GD
  - Incremental gradient descent
  - Performs parameter update for *each* training example and label (x and y)

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)})$$

# Gradient Descent

- Mini-Batch GD
  - Best of both worlds; performs update for every mini-batch of  $n$  training examples

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i:i+n)}; y^{(i:i+n)})$$



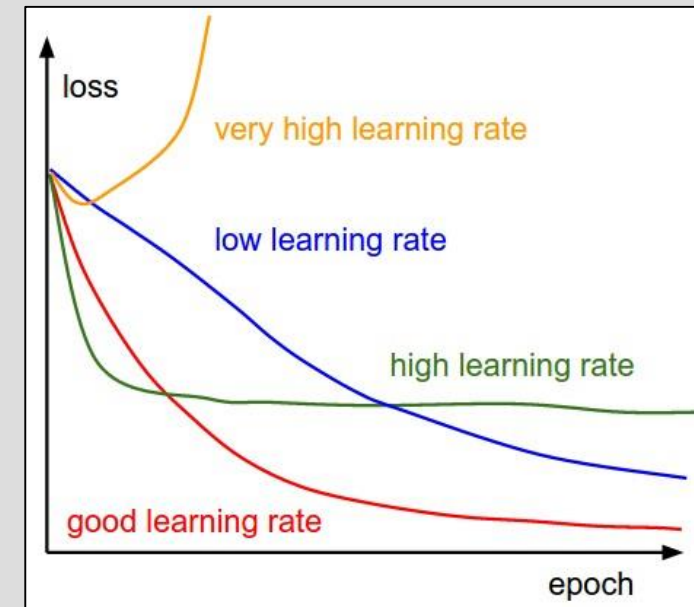
# Momentum and Learning Rate

$$\Delta w_{ij} = \left( \eta * \frac{\partial E}{\partial w_{ij}} \right)$$

weight increment      learning rate      weight gradient

$$\Delta w_{ij} = \left( \eta * \frac{\partial E}{\partial w_{ij}} \right) + (\gamma * \Delta w_{ij}^{t-1})$$

momentum factor      weight increment, previous iteration



# Questions?

