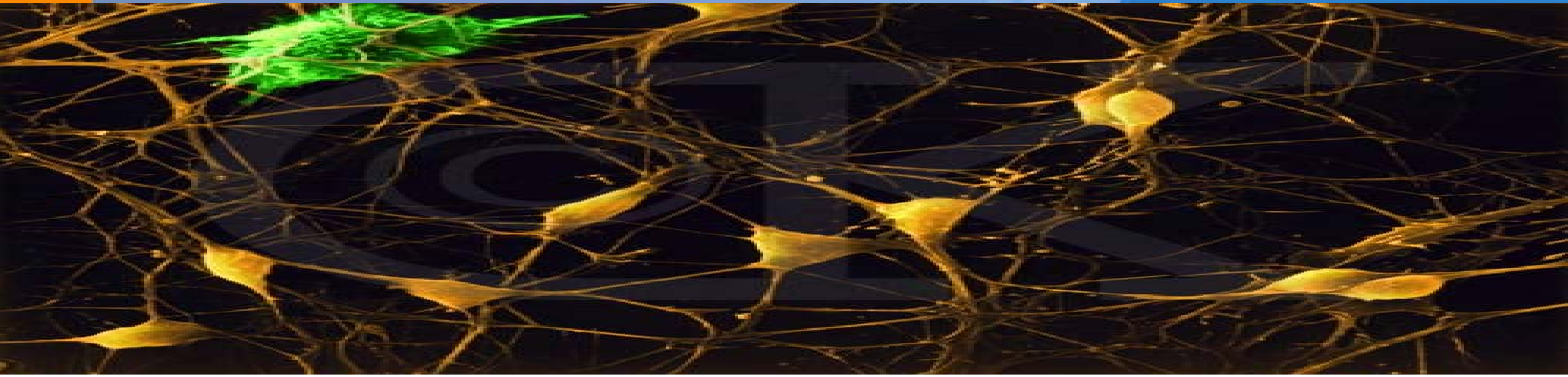
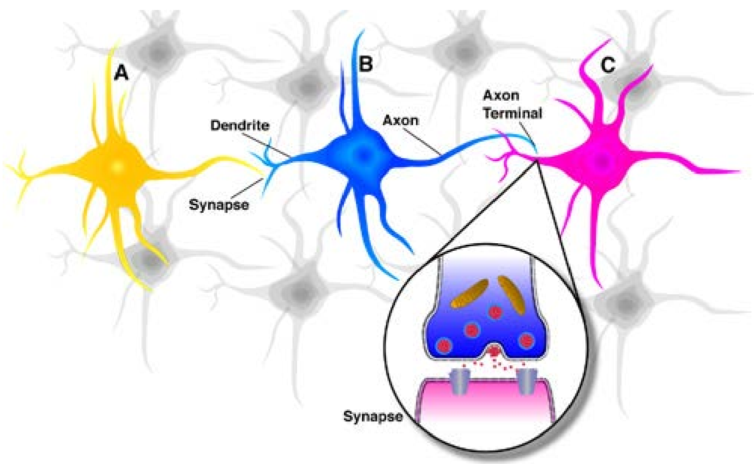


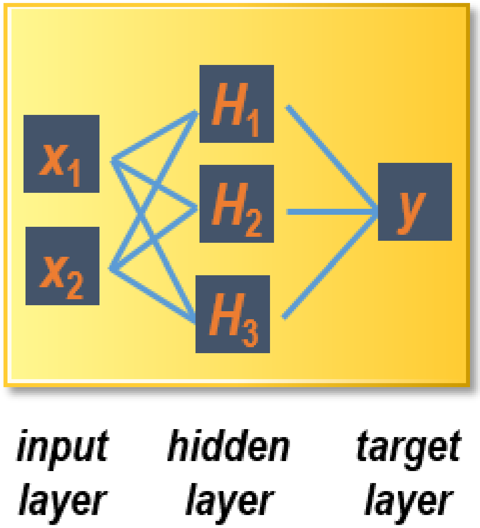
Artificial Neural Network

Dr. Pei Xu
Auburn University
10/31/2019

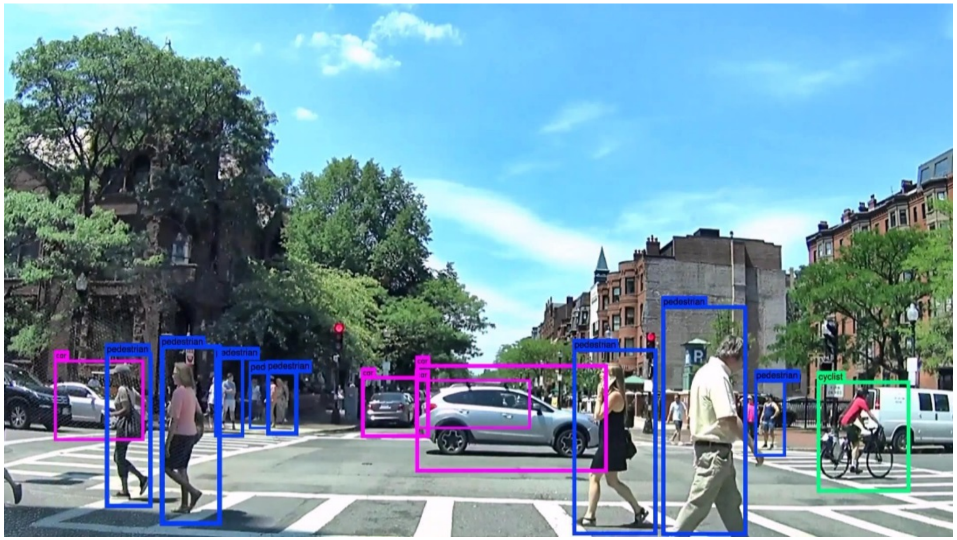
Artificial Neural Network

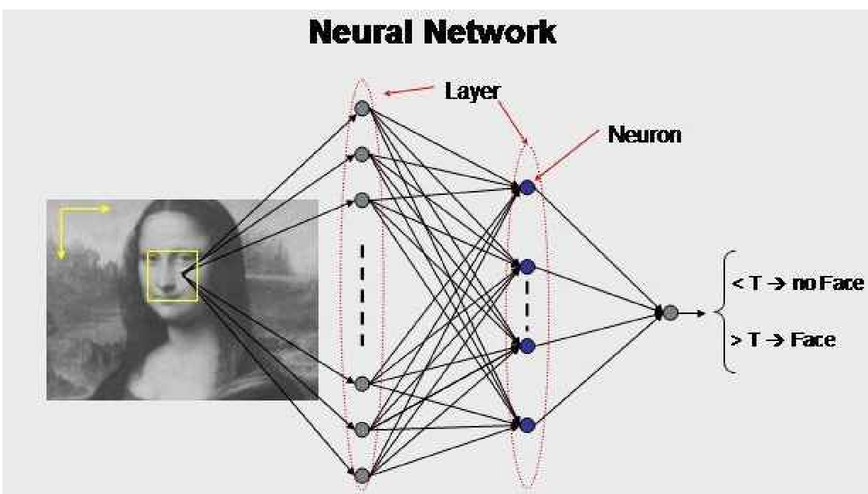
- An Artificial Neural Network is an information processing paradigm that is inspired by the biological nervous systems, such as the human brain's information processing mechanism.
- Each neuron mimics its biological counterpart
  - Taking various inputs
  - Sum up the inputs
  - Produce an output





Neural Network Applications





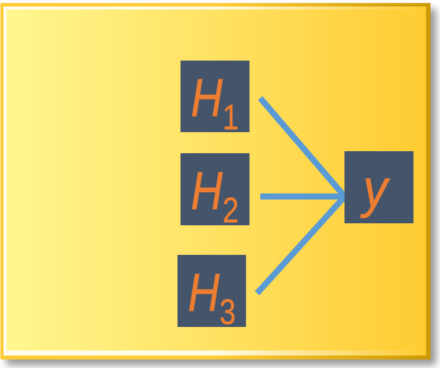
[Source]
Upper left: [Self Driving Car]
Upper right: [Cooking recipes]
Lower left: [Image Recognition]

Hidden Layer to Target Layer

hidden unit

$$\hat{y} = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$

bias estimate *weight estimate*



hidden layer *target layer*

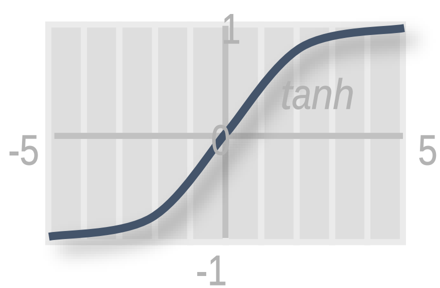
Input to Hidden Layer

hidden unit

$$\hat{y} = \hat{w}_{00} + \hat{w}_{01} \cdot H_1 + \hat{w}_{02} \cdot H_2 + \hat{w}_{03} \cdot H_3$$

bias estimate

weight estimate



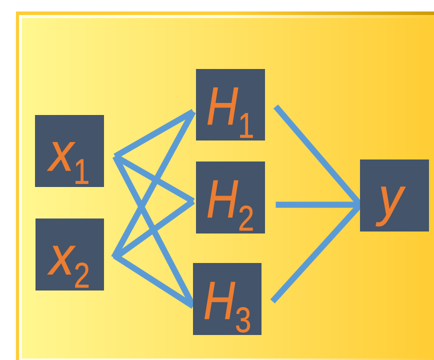
$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11}\, x_1 + \hat{w}_{12}\, x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21}\, x_1 + \hat{w}_{22}\, x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31}\, x_1 + \hat{w}_{32}\, x_2)$$

activation function

How many parameters need to be estimated in this model?



input layer

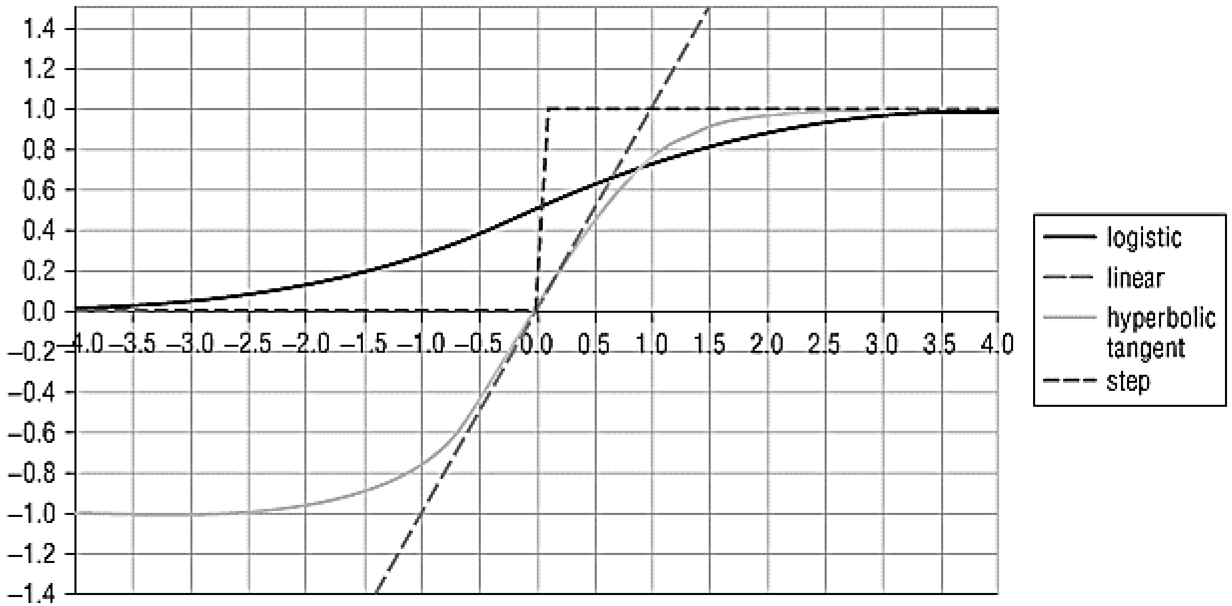
hidden layer

target layer

Activation Functions

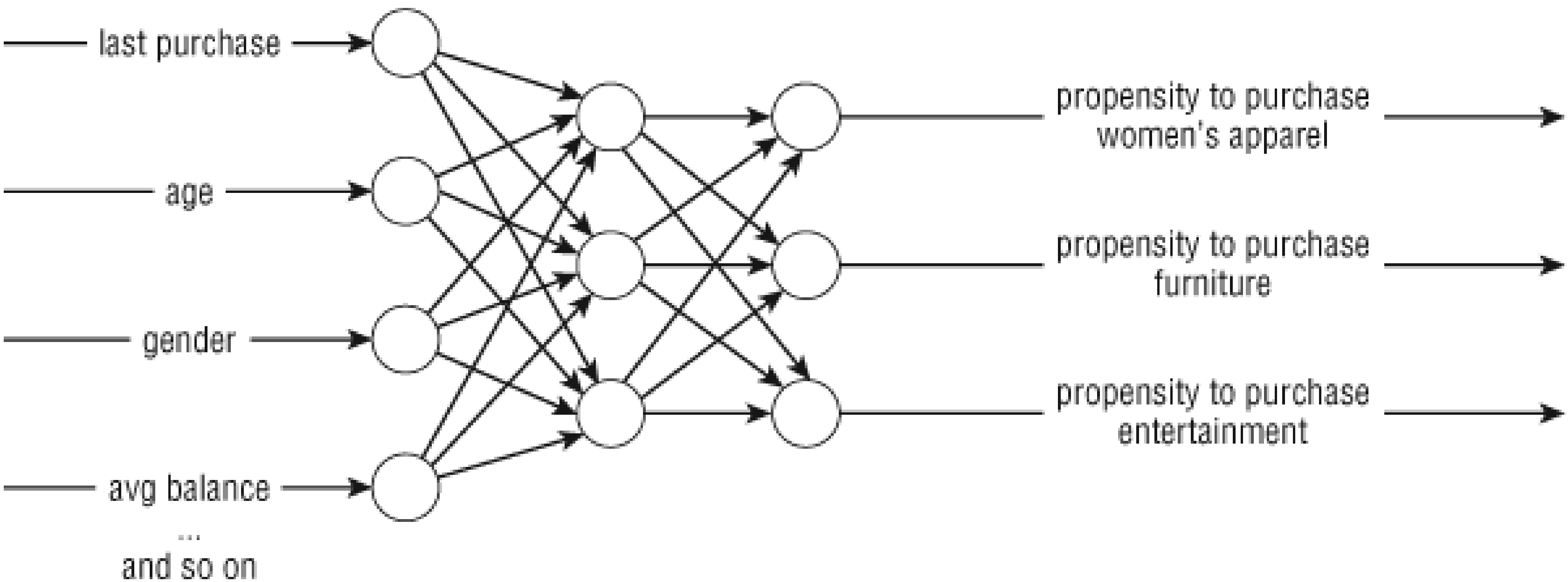
- The activation function is generally split into two parts
  - The *combination function* combines the inputs into a single value
  - The *transfer function* is used to produce an output of the hidden unit
    - The *hyperbolic tangent* is often preferred

      It produces values between -1 and 1, thus it spans both positive and negative values, and fall within a small range
    - When all inputs have similar ranges prevents one set of inputs from dominating other inputs



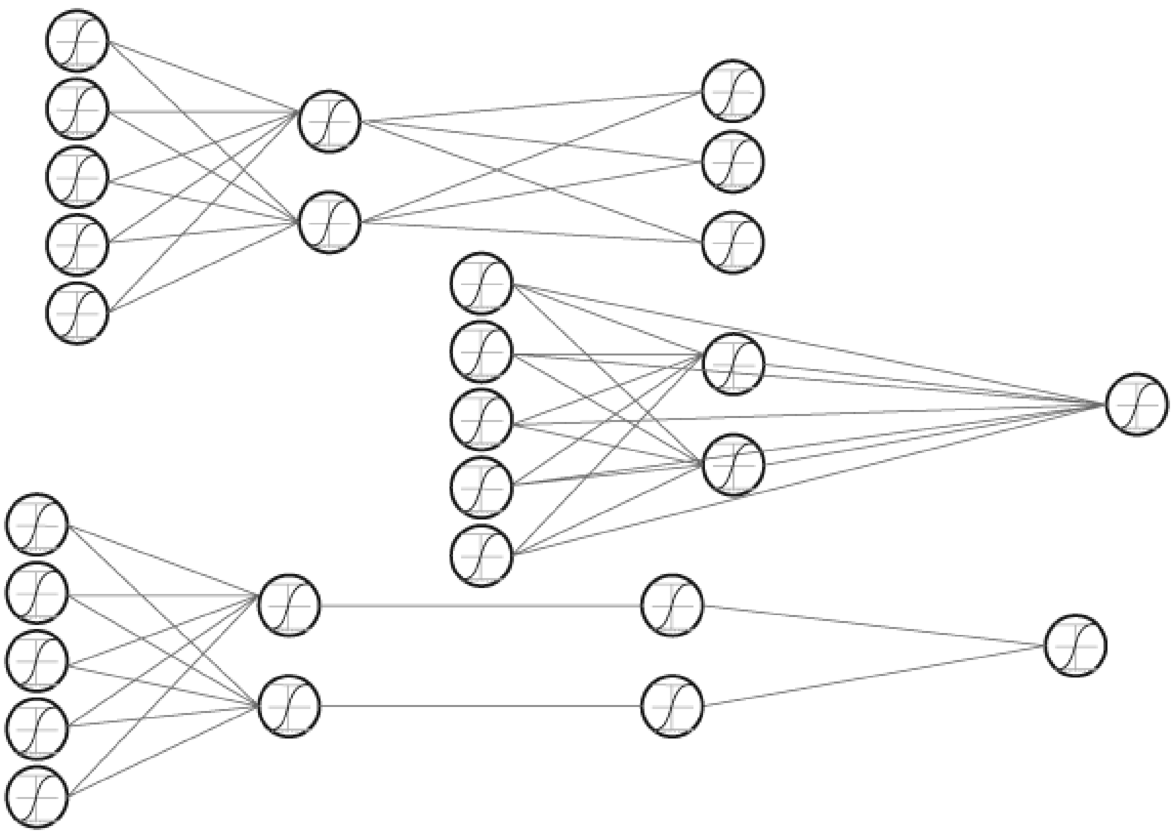
Network Topologies

- The output layer can have more than one unit
- - For example, a department store chain wants too predict the likelihood that customers will be purchasing products from various department.



Network Topologies

- There are several variations on the basic neural network architecture



Multi-layer perceptron (MLP)

- The most common network for predictive modeling is the *Multi-layer perceptron* (MLP)
  - has any number of inputs
  - has one hidden layer with any number of units
  - uses linear combination in the hidden and output layers
  - uses activation functions in the hidden layers
  - has any number of outputs with any activation function

Gradient Descent for estimating parameters

- Gradient Descent is a general approach to estimate parameters.
- It is similar to Ordinary Least Square (OLS) estimator in linear regression, but Gradient Descent can also be used in other models.

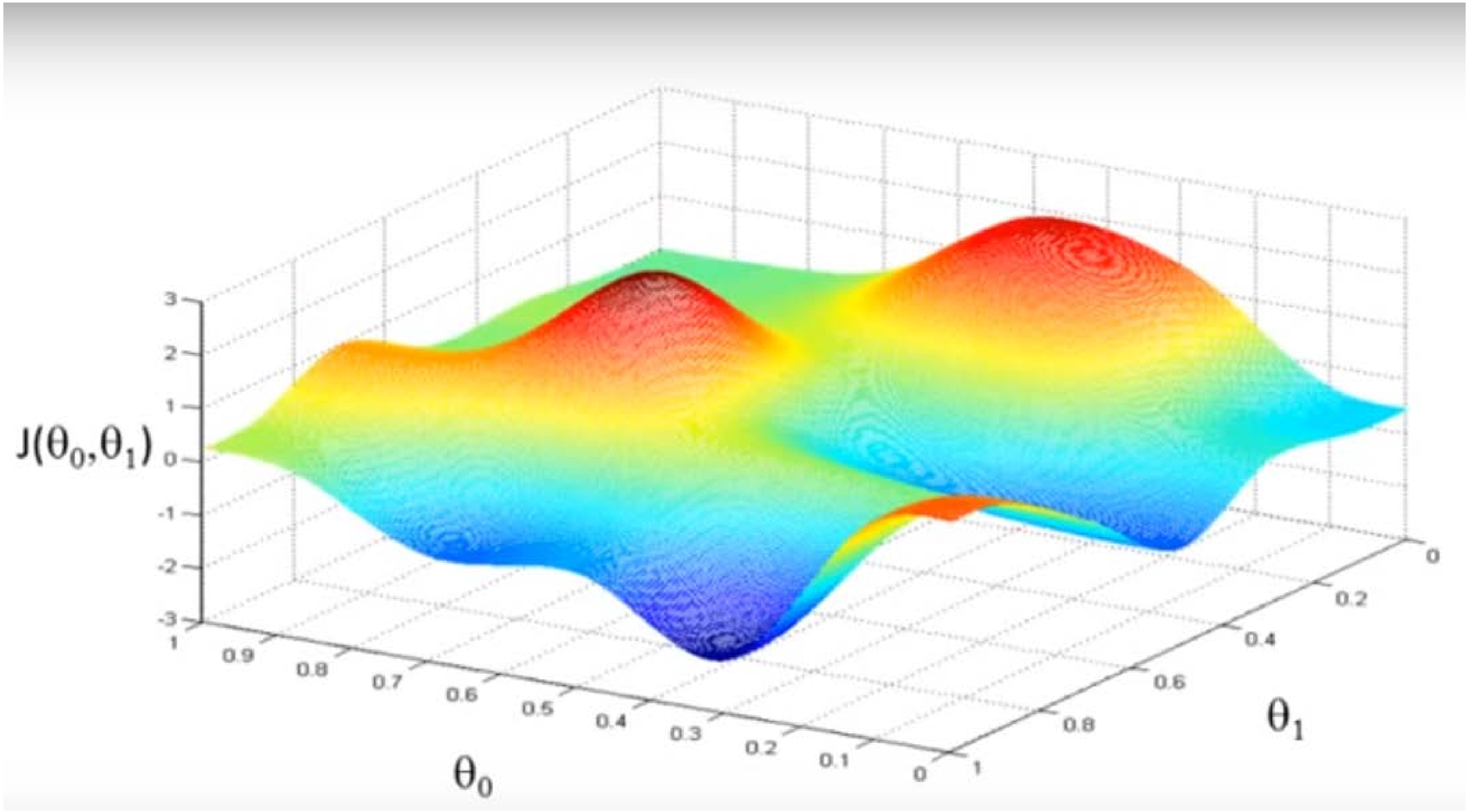
Have some function $J(\theta_0, \theta_1)$

Want $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Outline:

- Start with some θ_0, θ_1
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$ until we hopefully end up at a minimum

Gradient Descent – Local optimal



Gradient descent algorithm

repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \qquad (\text{for } j = 0 \text{ and } j = 1)$$

}

Correct: Simultaneous update

$$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$
$$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$
$$\theta_0 := \text{temp0}$$
$$\theta_1 := \text{temp1}$$

Strength

- Neural Networks have outperformed other approaches -- and not by a little, by a lot - on some of the most challenging tasks in the history of artificial
  - image recognition, image captioning, natural language processing
- Like Decision Tree, Neural Network can almost approximate any kind of patterns.

Drawbacks

1. Requires huge amounts of data
2. Variable selection is an issue

   It works best when there are only a handful of input variables, but the technique itself does not help choose which variable to use
3. Working out the topology of the neural network for a problem is a "black art“
4. No guarantee that the resulting set of weights is optimal
5. Cannot explain what it is doing

- Overall, Neural network is best for problems where a black box model is acceptable.

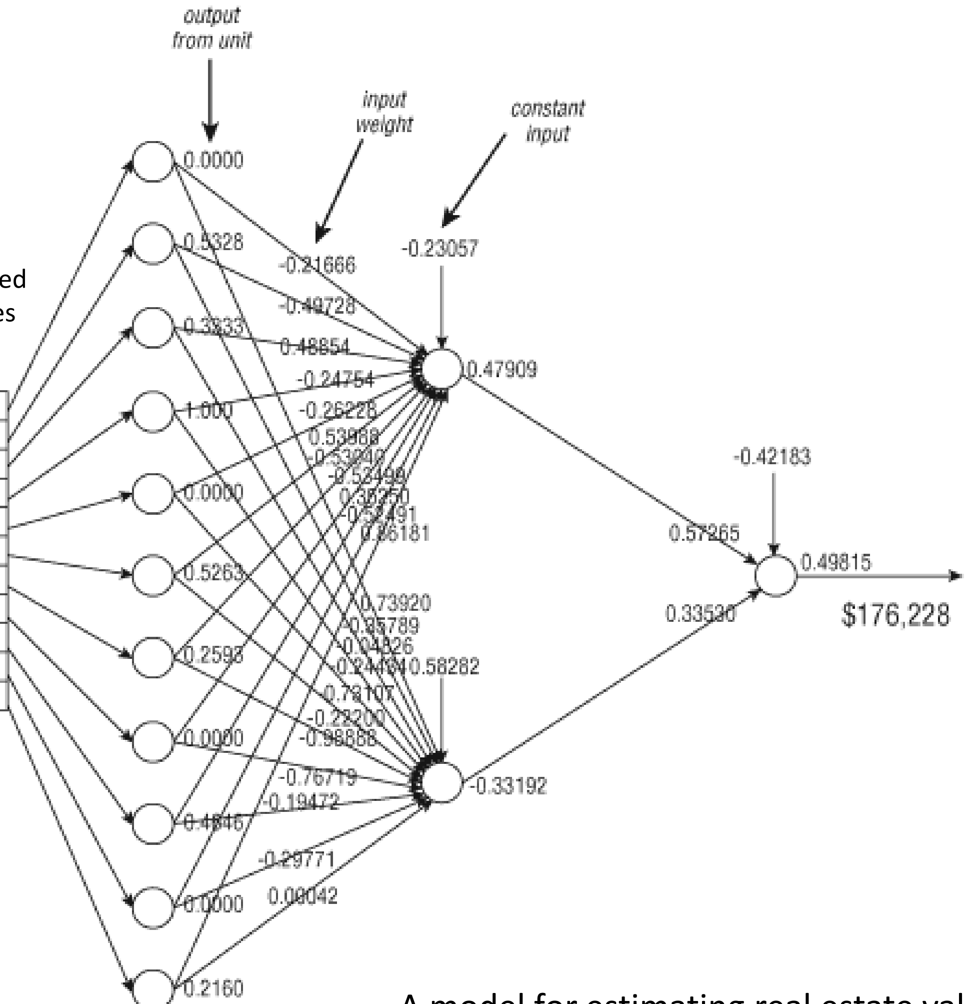
Sensitivity Analysis

- Set all predictor values to their mean and obtaining the network's prediction.
- The process is repeated by setting each predictor sequentially to its minimum, and then the maximum, value.
- By comparing the predictions from different levels of the predictors, we can get a sense of which predictors affect predictions more and in what way.

A Network Example

Standardized input values

Num_Apartments	1	0.0000
Year_Built	1923	0.5328
Plumbing_Fixtures	9	0.3333
Heating_Type	B	1.0000
Basement_Garage	0	0.0000
Attached_Garage	120	0.5263
Living_Area	1614	0.2593
Deck_Area	0	0.0000
Porch_Area	210	0.4646
Recroom_Area	0	0.0000
Basement_Area	175	0.2160



A model for estimating real estate values

Review Questions

- ***Q1**. The importance of an input variable in predicting a target in an MLP-based neural network can be figured out by which of the following?*

  a. the highest absolute value of the parameter estimate between the input and any of the hidden neurons

  b. the average of the absolute values of parameter estimates between the input and all of the hidden neurons

  c. the highest absolute value of the parameter estimate between the input and any of the hidden neurons multiplied by the absolute value of the parameter estimate of the hidden neuron

  d. none of the above

Review Questions

- ***Q2.*** *In preparation for a neural network model, is imputation of missing values needed? Why or why not?*

- ***Q3.*** *In preparation for a neural network model, is data transformation generally needed? Why or why not?*

Review Questions

- ***Q2.*** *In preparation for a neural network model, is imputation of missing values needed? Why or why not?*

  Yes. Neural network models, as well as most models relying on a prediction formula, require a complete record for both modeling and scoring.

- ***Q3.*** *In preparation for a neural network model, is data transformation generally needed? Why or why not?*

  Not necessarily. Neural network models create transformations of inputs for use in a regression-like model. However, having input distributions with low skewness and kurtosis tends to result in more stable models.

Reference

- Textbook. Introduction to Data Mining. Chapter 4.
- Berry, Michael J., and Gordon Linoff. *Data mining techniques: for marketing, sales, and customer support* (3rd Edition). John Wiley & Sons, Inc., 2011.
- “Gradient Descent in Machine Learning” By Andrew Ng, 2014

  https://www.youtube.com/watch?v=P3K38HusyV4