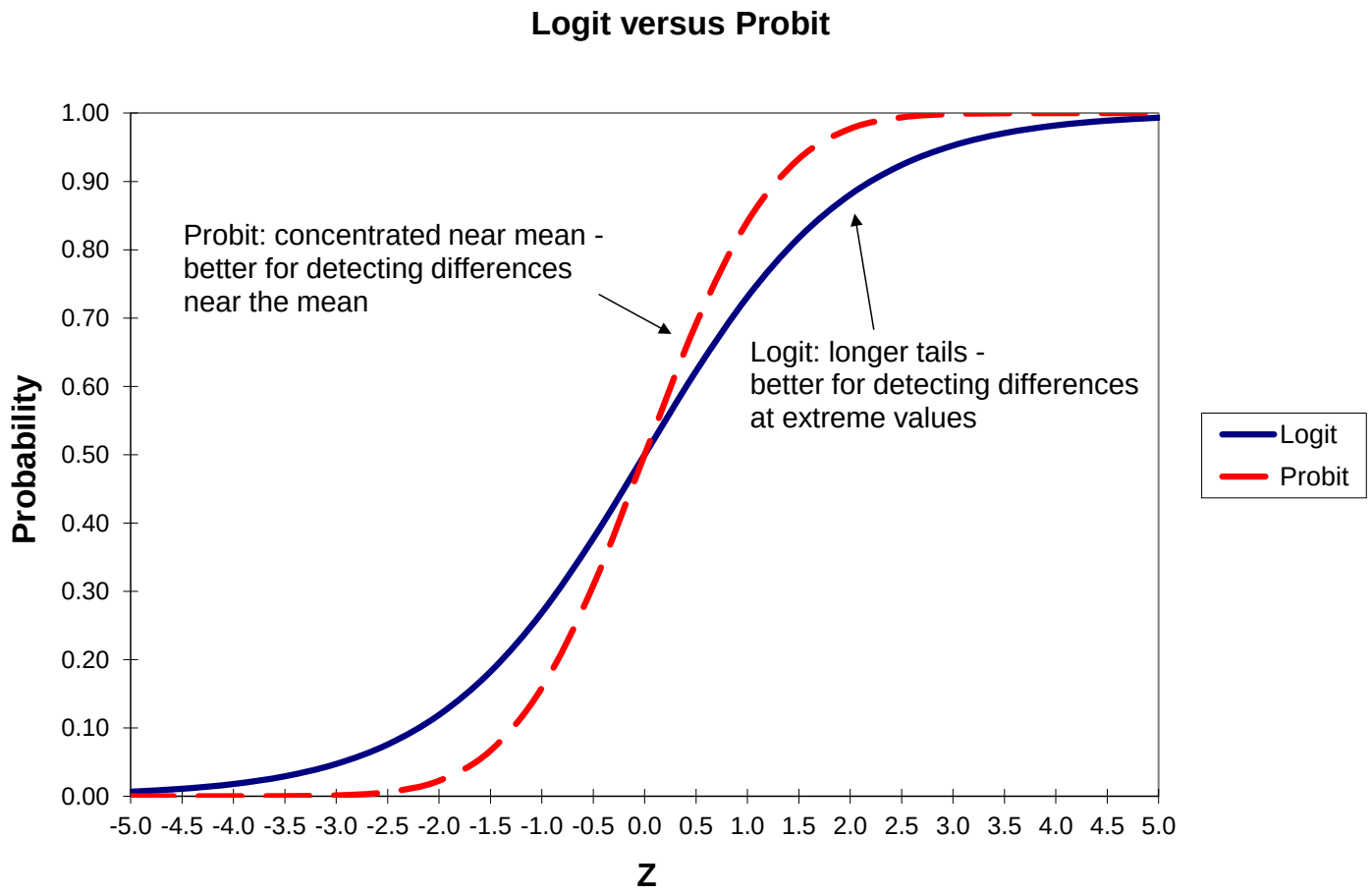
Choice Models Overview

Choice Models Overview

- Choice models have Y-variables that are zero or one, e.g., when a customer enters a store, do they buy a TV or not?
- Some techniques to represent choice models include logit, probit, and neural networks.
- Logit and probit have Y-variables which are binary (zero or one) and predict probabilities.
- Logit and probit have s-shaped curves (monotonic): always increasing or always decreasing.
- Neural networks are not limited to monotonic behavior or binary Y-variables.

Choice Models Overview (cont.)

- Logit uses the logistic distribution and is more sensitive to extreme values of X.
- Probit uses the normal distribution and is more sensitive near the mean of X.

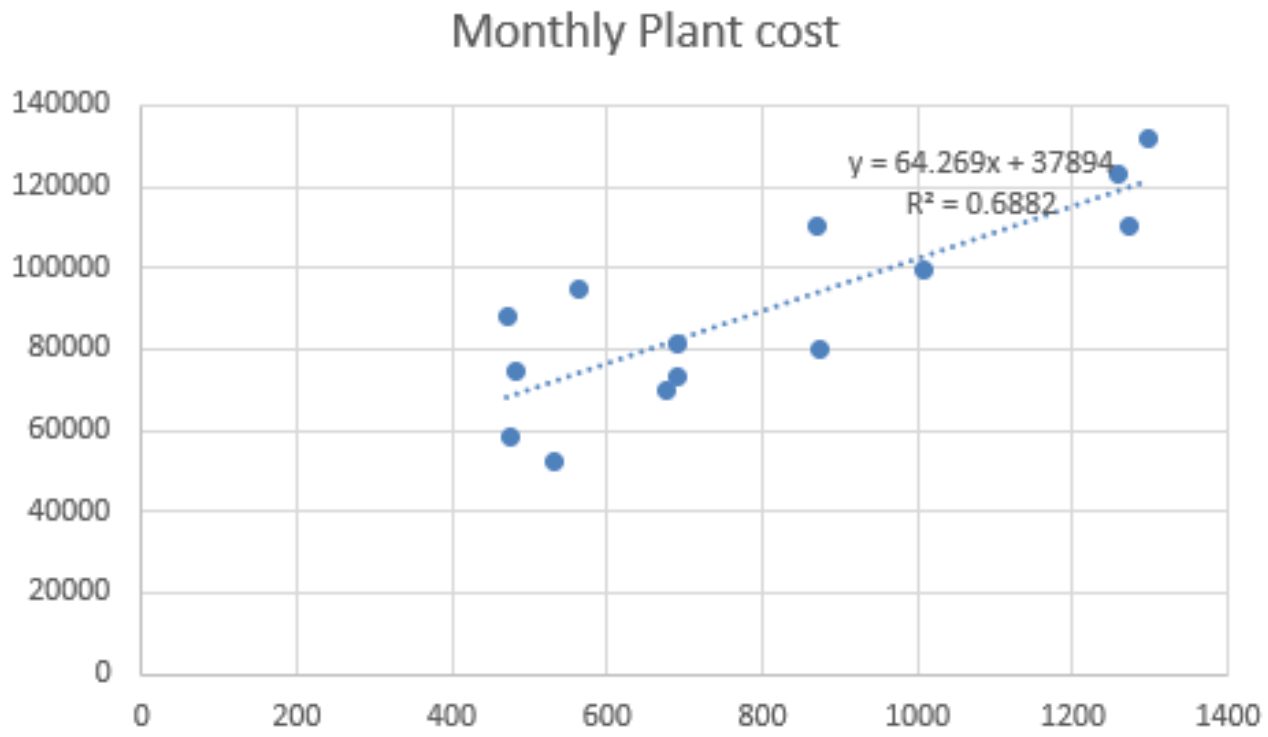


The End

Review of Linear Regression

Review of Linear Regression

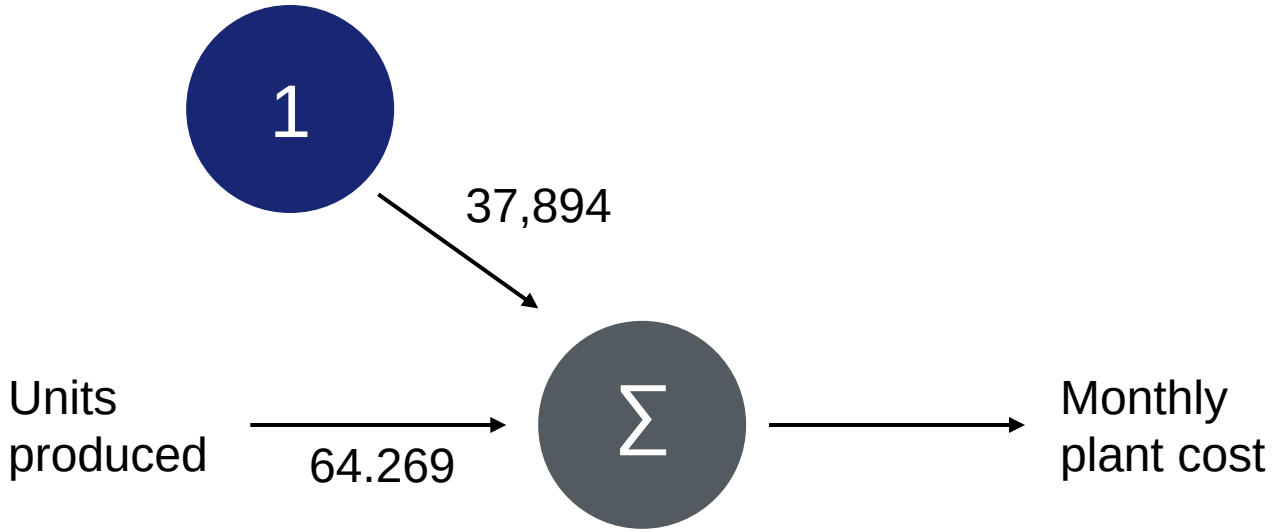
Linear regression finds the line that best fits the data by minimizing the square of the error terms (residuals)



Review of Linear Regression (cont.)

The linear regression equation can be represented as:

Monthly plant cost = 37,894 + 64.269 * Units produced

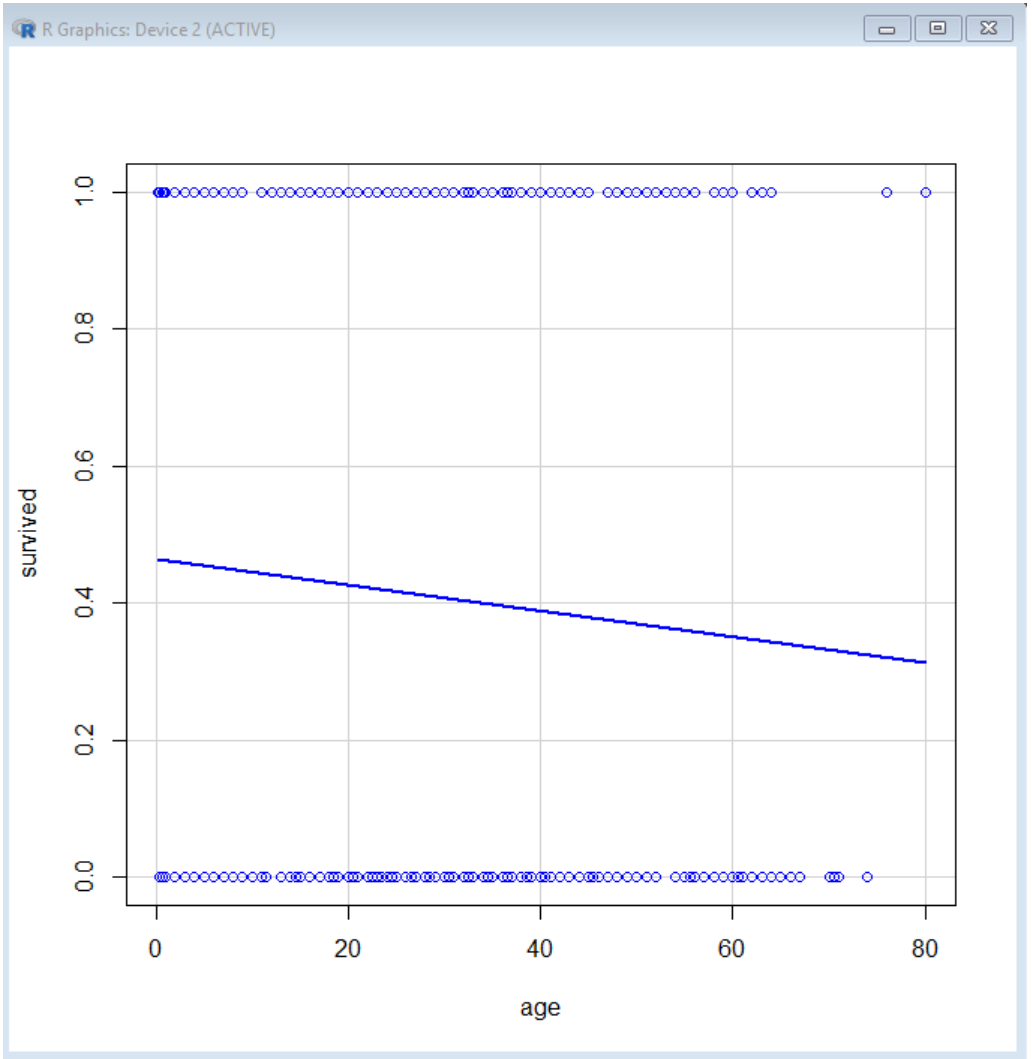


The End

Review of Logit and the Logistic Function

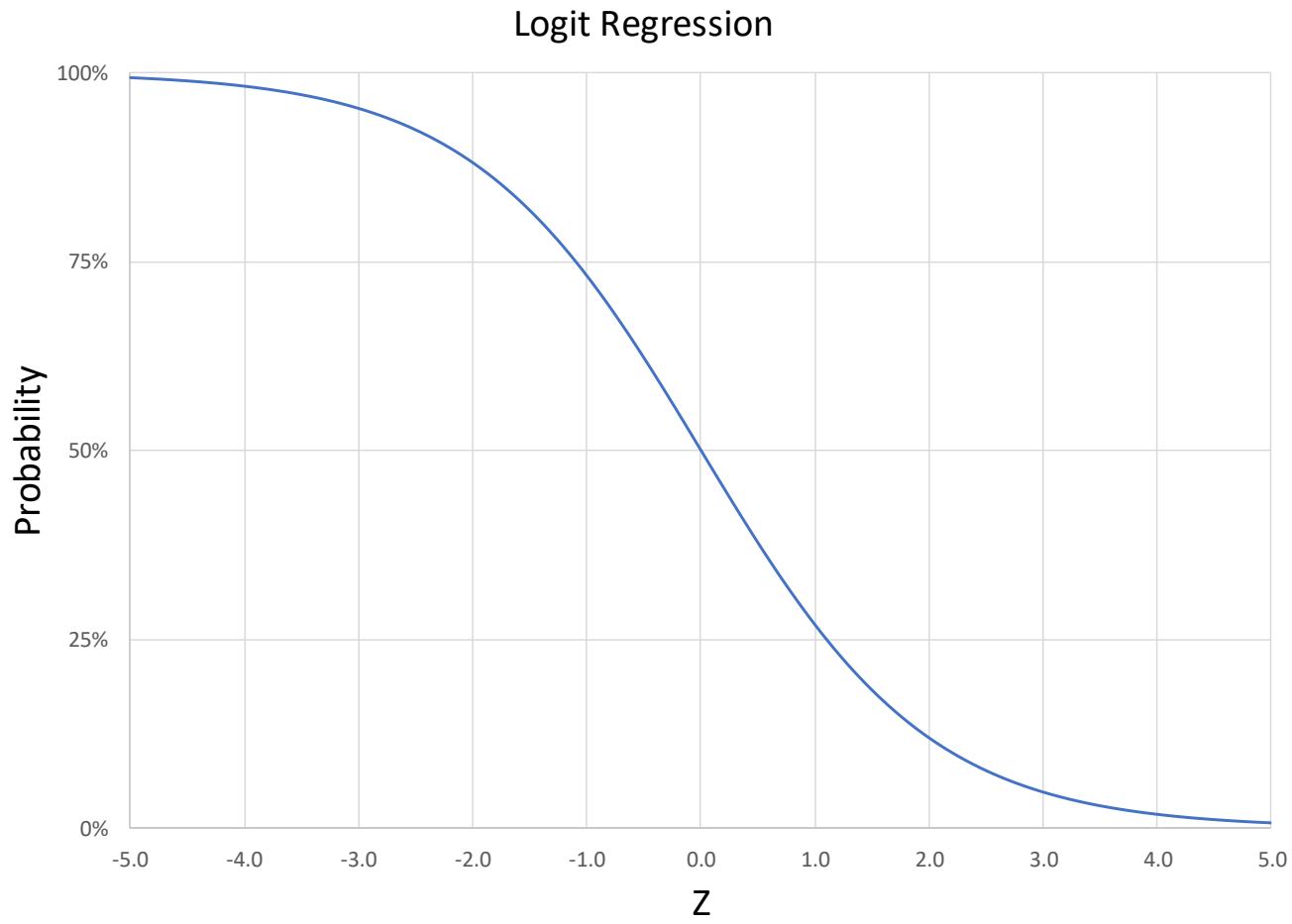
Review of Logit and the Logistic Function, Part I

Linear regression does not work when the Y-variable is binary (zero or one)



Review of Logit and the Logistic Function, Part II

For binary Y-variables, logit creates an s-shaped curve using the logistic function



Review of Logit and the Logistic Function, Part III

The logistic function is

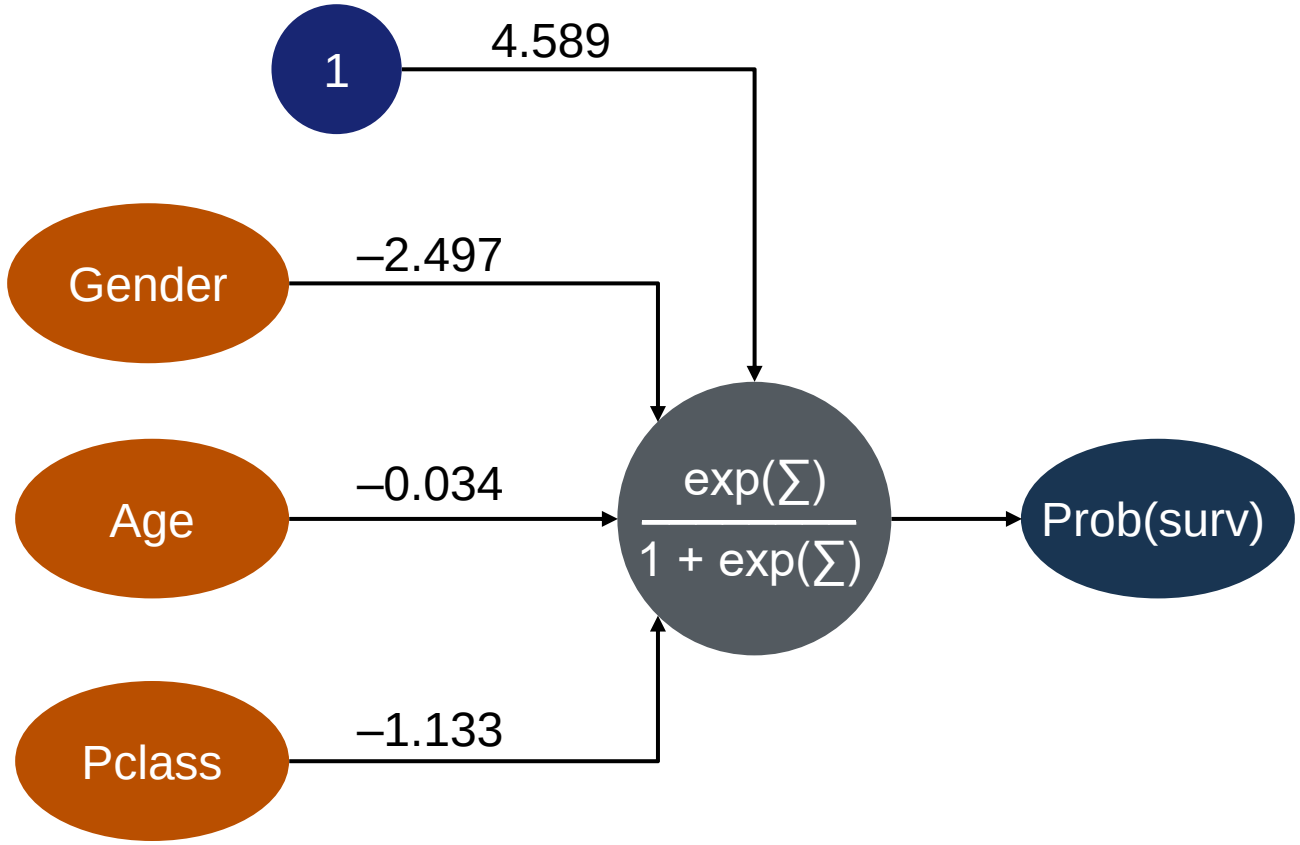
$$f(X) = \exp(\sum \beta_i X_i) / (1 + \exp(\sum \beta_i X_i))$$

where:

- β_i = the coefficients in a logit regression
- X_i = the variables in a logit regression

Review of Logit and the Logistic Function, Part IV

Pictorially, this looks like:



The End

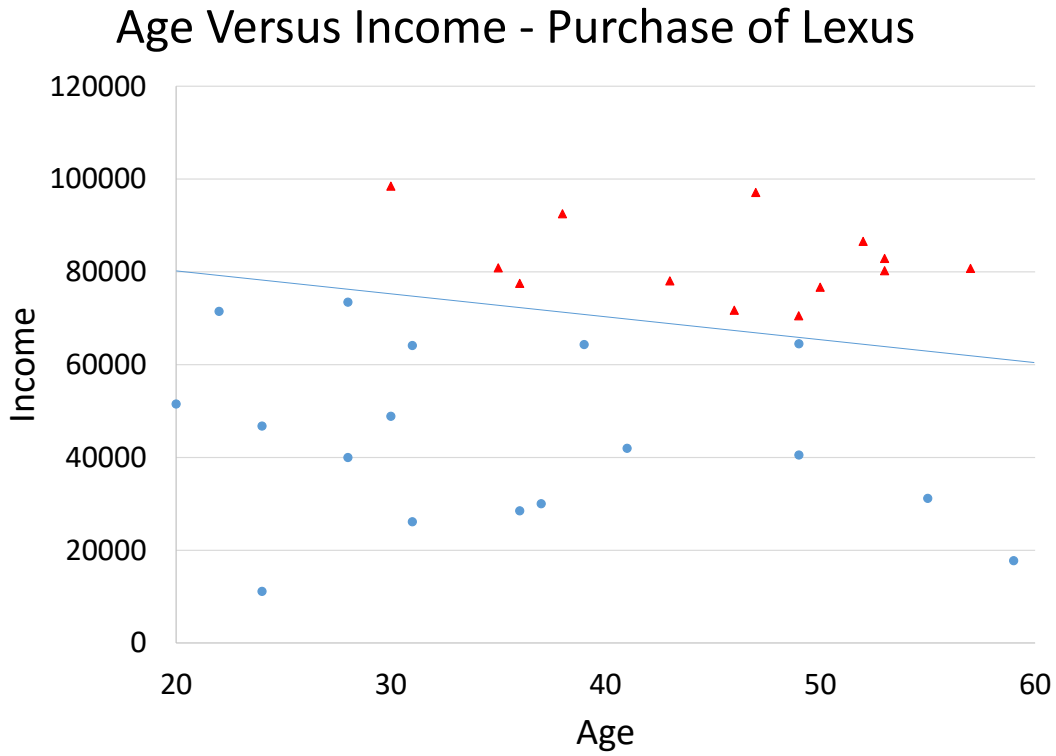
Perceptrons

Perceptrons, Part I

- In 1957, Frank Rosenblatt created the perceptron learning algorithm.
- Perceptrons were linear classifiers, i.e., they separated groups which had Y-variables which were zero or one.

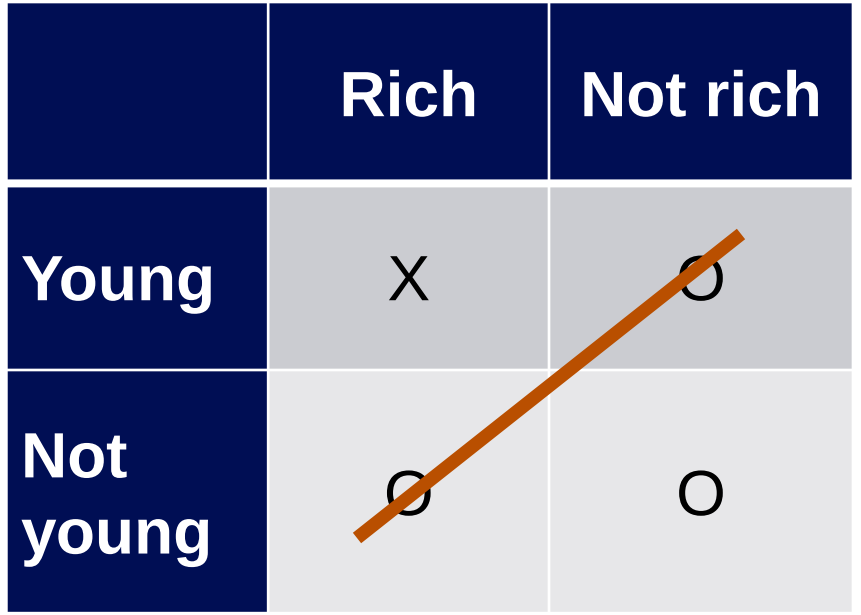
Perceptrons, Part II

- In this example, datapoints above the line represent people who bought a Lexus, those below the line did not.
- Buyers tend to have higher income or are slightly older.
- The perceptron identifies the straight line which separates the two groups.



Perceptrons, Part III

- In 1969, Minsky and Papert analyzed the effectiveness of perceptrons.
- Perceptrons worked on AND conditions; in this example, X identifies those who are both rich and young.
- The perceptron separates the rich and young from the others.



Perceptrons, Part IV

- Perceptrons worked on OR conditions; in this example, X identifies those who are rich or young.
- The perceptron separates the rich or young from the others.

	Rich	Not rich
Young	X	X
Not young	X	O

Perceptrons, Part V

- Perceptrons did not work on the exclusive OR (XOR).
- Exclusive OR means rich or young, but not both.
- In the 1970s, interest in perceptrons died because of the XOR problem.
- No straight line could be drawn to separate the groups.

	Rich	Not rich
Young	O	X
Not young	X	O

The End

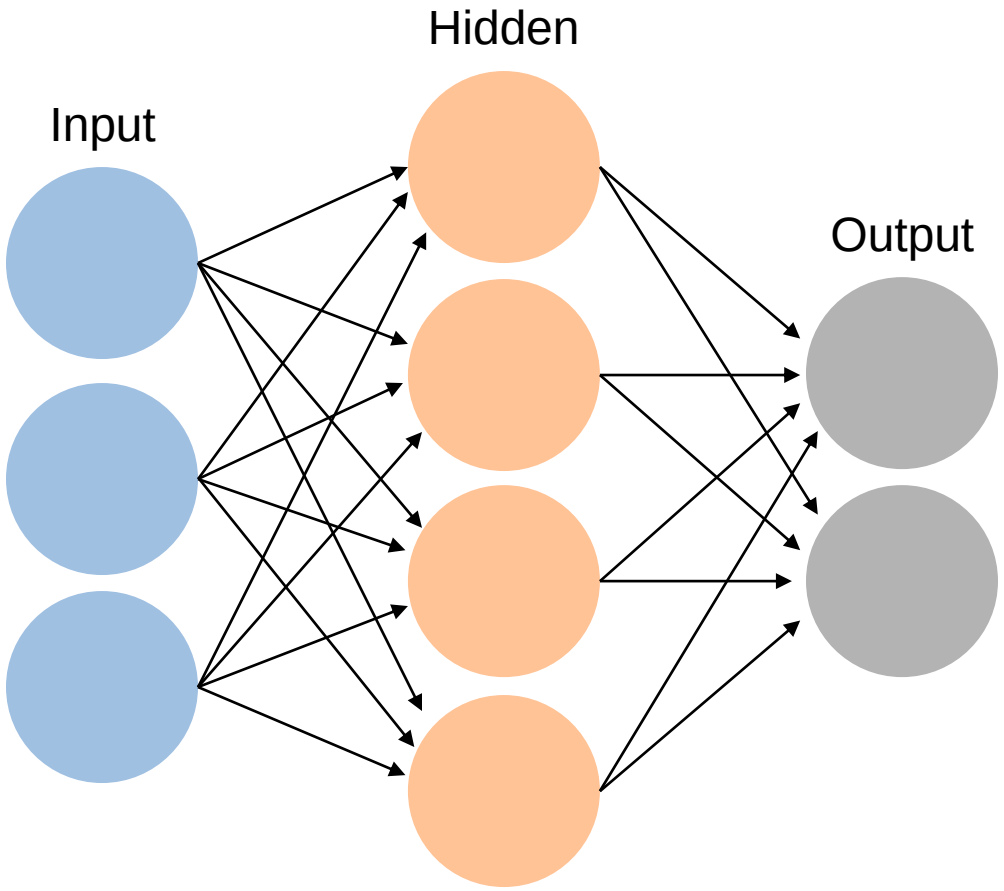
Neural Networks

Neural Networks, Part I

- In the 1980s, researchers revisited perceptrons, with a twist
- Instead of using linear separators (straight line), they changed to curves using the logistic function
- In addition, instead of having inputs determine outputs, they introduced a middle layer called the hidden layer
- Inputs determined the values of nodes in the hidden layer; nodes in the hidden layer determined the values of the output
- The hidden nodes and the output use the logistic function

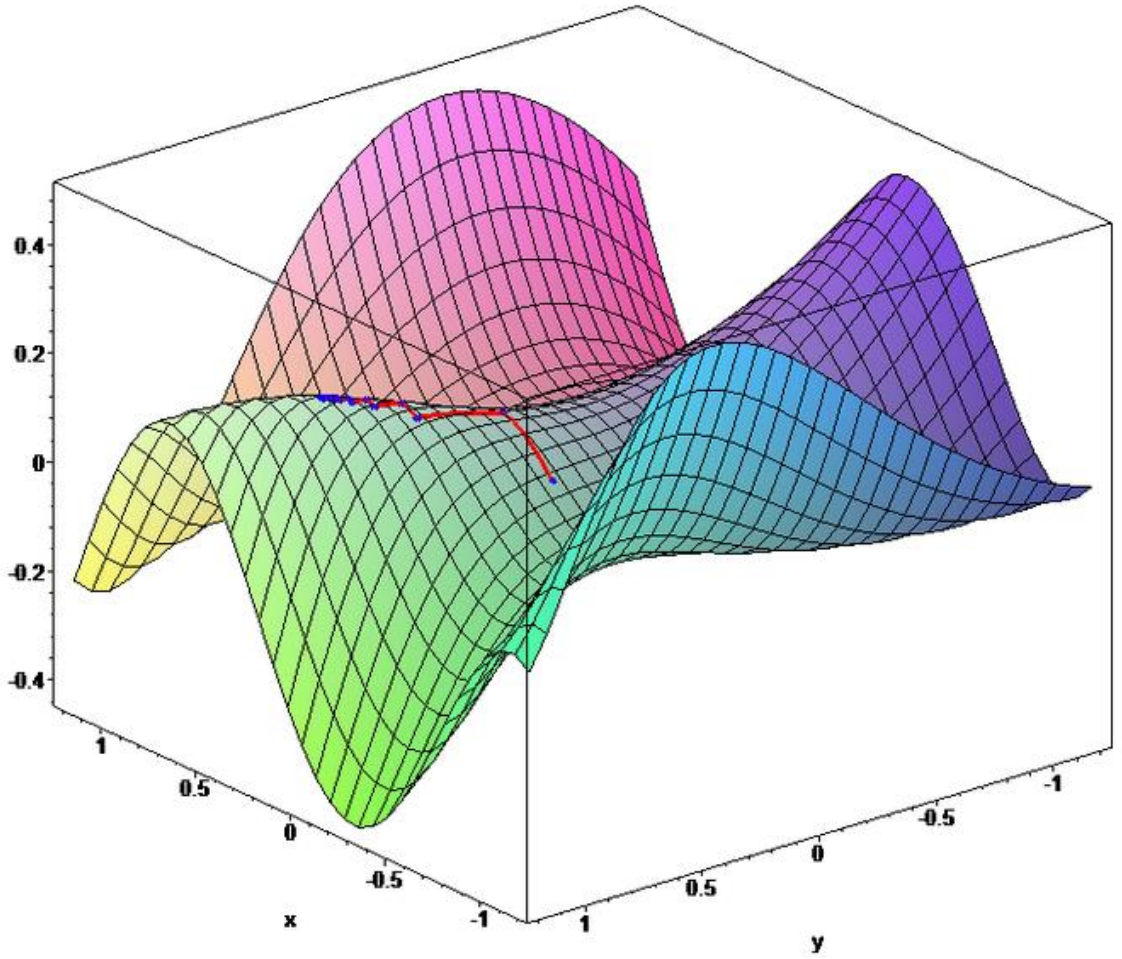
Neural Networks, Part II

- Each hidden node and each output uses a logistic function



Neural Networks, Part III

- The back propagation algorithm uses gradient search (hill climbing) to find the best solution
- It randomly selects a starting point, then calculates the slope and climbs the hill
- When there are multiple hills, it sometimes get stuck in a local optimum



Neural Networks, Part IV

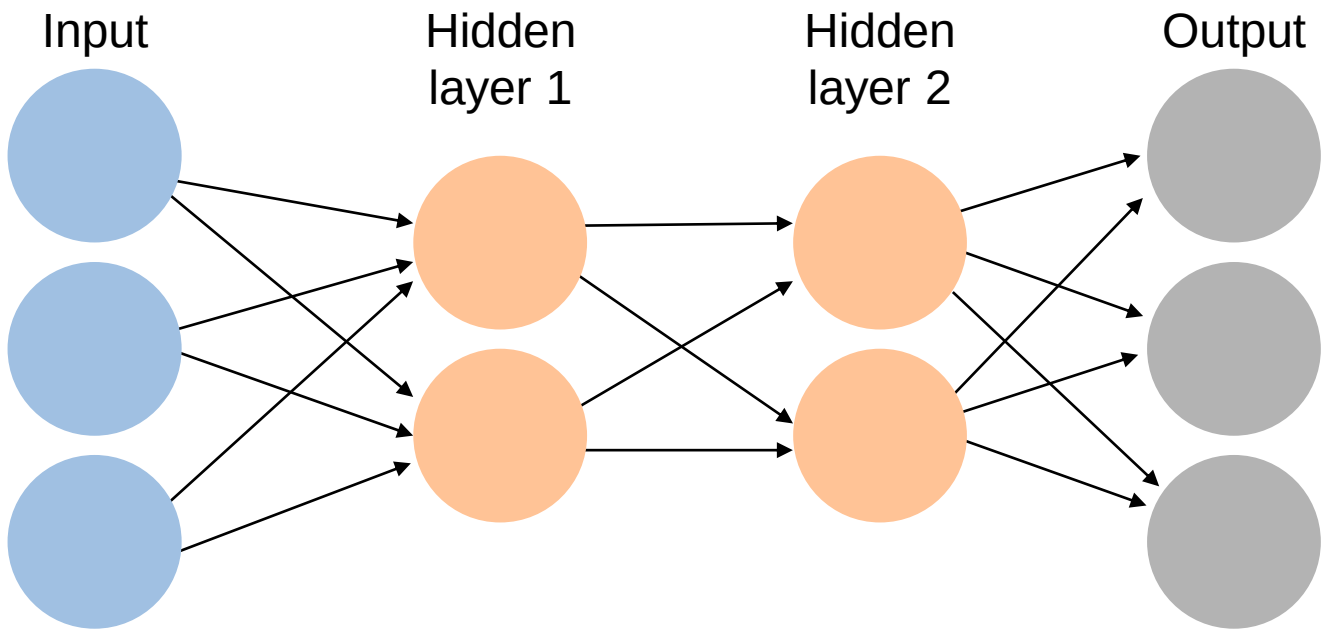
- Advantages
  - Can have binary or linear outputs
  - Can have one or more outputs
  - Can represent any mathematical relationship
  - Have been used in voice recognition and speech generation (SIRI), photo recognition (law enforcement), and self-driving cars (Tesla)
  - Fast to make decisions after they have learned
- Disadvantages
  - Can get stuck in local optima; need to run several times
  - Slow to learn

The End

Deep Learning

Deep Learning

- Deep learning neural networks have two or more hidden layers.
- Deep learning has been used in sophisticated game analysis, such as the game of Go.



The End