Neural networks = Layers of nodes

Each node has an activation variable (continuous value from 0 – 1)

Activations of an output layer are completely determined by activation of input layer

Biomimicry – take the weighted sum of input + bias and apply activation function

Weighted edges between one node and the next

Dense input and output layer is equivalent to complete bipartite graph

Data structure

Inspired by biology, but its not the same

e.g. activations of neurons are continuous variable where in the brain we observe a all

or-none response

How do neural networks (functions) “learn”?

Learning = Making more accurate predictions = Minimizing error in predictions

Suppose you have a set of input values and output values (training patterns)

Across the set of training patterns, you can calculate SSE (Sum-squared error)

Find the Gradient of SSE (based on changes in weights and biases)

Change weights in direction opposite of gradient (descending the cost function towards

A local minima