# Neural Networks

A neural network consists of neurons which are named after neurons in the human body. Neurons receive impulses (features) and have some resistance (weight). In computers, a neuron consists of an adder which sums weights and features:

$$\sum_i w_i x_i$$

and also consists of an activation function ($g(x)$). This differentiates it from a linear regression algorithm. Neural networks can have more than one neuron in which inputs with their own respective weights can go into.

Input layers consist of inputs and their weights. Hidden layers consist of neurons, each with their own adder and activation function. Neural networks can have multiple outputs in the output layer.

# Activation Functions

---

Denoted as $g(x)$, the activation function takes the added weights and prepares it for output. A linear activation function returns what is inputed: $g(x) = x$. This is what the single neuron linear regression network.

# Neural Network Architecture

---

- Input Layer
- Hidden Layer
- Output Layer

Each input in the input layer is an input to each neuron in the first hidden layer. The subsequent neurons are themselves inputs into each neuron into the next layer. The output layer is the output if the final hidden layer. Each input has a specific weight attached to it.

## Logistic Regression

Logistic regression is like linear regression but is used for classification. It's a classification algorithm that can be referred to a Log-odds ratio or Logit:

$$\frac{p}{1-p}$$

If the Log-odds ratio is $\log(\frac{p}{1-p})$, how can we integrated it with linear regression?

$$\log_e \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

If we take the partial derivative with respect to some weights, we can determine the best activation function for a network. To prove the log-odds function:

$$p = \frac{e^{x\beta}}{1+e^{x\beta}} = \frac{1}{1+e^{-x\beta}}$$

The final result is easier to derive when it comes time for other linear regression functions. The derivative results to:

$$\frac{d}{dx}[\frac{1}{1+e^{-x}}] = \frac{e^{-x}}{(1+e^{-x})^2}$$

Therefore, to take the derivative of an activation function, it is simply

$$g'(x) = g(x) \cdot (1 - g(x))$$