Introduction to neural networks: Takeaways

Syntax

- Creating a histogram for every numeric column in a DataFrame:

```
df.hist()
```

- Defining the sigmoid activation function:

```
def sigmoid_activation(x, theta):
x = np.asarray(x)
theta = np.asarray(theta)
return 1 / (1 + np.exp(-np.dot(theta.T, x)))
```

Concepts

- Neural networks are a class of models that can learn from non-linear interactions between variables.
- Neural networks are very loosely inspired by the structure of neurons in the human brain. The models are built by a series of activation units, known as neurons, to make predictions of some outcome. Neurons take in some input, apply a transformation function, and return an output.
- You can use an activation function to determine the output of a neural network. You can use the popular sigmoid activation function because it returns values between 0 and 1 and can be treated as probabilities.
- The sigmoid function is defined as the following:

$$g(z) = \frac{1}{1 + e^{-x}}$$

- The sigmoid function is a key component of the popular sigmoid activation function:

$$h_\Theta(x) = \frac{1}{1 + e^{-\Theta^T x}} = \frac{1}{1 + e^{-(\theta_0 1 + \theta_1 x_1 + \theta_2 x_2)}}$$

- In gradient descent, we need a minimize a cost function, which measures the error in our model. The cost function measures the difference between the desired output and the actual output, defined as:

$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^{m} (y_i * log(h_\Theta(x_i)) + (1 - y_i) log(1 - h_\Theta(x_i)))$$

- Calculating derivatives are more complicated in neural networks than in linear regression. We can compute the derivative using the chain rule:

$$\frac{\partial J}{\partial \theta_j} = \frac{\partial J}{\partial h(\Theta)} \frac{\partial h(\Theta)}{\partial \theta_j}$$

- Neural networks are usually built using multiple layers of neurons. Adding more layers to the network allows for more complex functions.
- Backpropagation focuses on updating parameters starting at the last layer and circling back through each layer.
- The AUC, or area under the curve, score of the receiver operating characteristic benchmarks how well a neural network performs.

Resources

- Neural Network
- Backpropagation
- How to Implement the Backpropagation Algorithm