**Supervised Learning**

* We have past data telling us what the correct answer is.
* The goal is to then predict the correct answer, based on new data, represented in the same way as the past data.

**Unsupervised Learning**

* We provide a ton of data and ask the model to create clusters of similarity/usefulness.
* There is no predefined answer, as there is with supervised learning.

**Classification**

* Predicting a discrete value, which could be one of n number of options
* Example, predicting if an object detected is metal, plastic, or another material

**Regression**

* Predicting a continuous value, which could be any number.
* Example, predicting the price of a house in a given area.

**Bias Variance Trade-off**

* Bias error is the difference between the predicted value and the correct value.
* Variance is how similar your bias is from one sample to another.
* Ideally, you want as little of each as possible, however, they are inversely related.
* To combat high bias, you could try to add more variables, or explore more complex algorithms.
* To combat high variance, you could try using a validation set, removing outliers through standardisation, or using regularisation techniques (a set of methods to reduce overfitting).

**Feature Scaling**

* Transforming raw values into some standard scale.
* This is done because gradient descent may struggle to converge without proper scaling
* Clustering algorithms will be skewed by differing scales as they rely on distance metrics
* There are two common types;
  + Min-Max normalisation. Typically set values to be between 0 and 1.
  + In this case, the largest number in your dataset will be represented as a 1 and the smallest number will be represented as a 0.

* + Standardisation. Representing values in standard deviations from the mean, usually in the ballpark of -3 to 3
  + X = Current value in the list we are transforming
  + = The average of all values in the original list

= The standard deviation of the original list