Week 5 task – write a short summary covering:

* Is it Supervised/Unsupervised/Reinforcement learning?
* What does the algorithm do and what situations will it be most useful?

**Linear regression**

Linear regression is an example of supervised learning. It is the process of finding the correlations between dependent and independent variables. Linear regression algorithms are used for predicting continuous values such as price, salary, age. It can be used in situations that need to predict market trends or house prices. You might also use this algorithm to estimate the impact of factors such as gender, years of education, and highest qualification held on wage.

**Logistic regression**

Similar to linear regression, logistic regression is an example of supervised learning and a way of finding the correlations between dependent and independent variables. Logistic regression algorithms are used for predicting response where the response variable is categorical in nature e.g. for a yes/no, true/false, red/green/blue, pass/fail type response. For example, you would use logistic regression to predict if based on size a house would sell for more than £200K. Rather than continuous the possible outputs are either Yes, the house will sell for more than £200K, or No, the house will not. It can also be used in detecting spam email (spam/not spam) and by banks in detecting credit card fraud (fraud/not fraud) based on variables such as amount spent, place and time.

**Decision Tree**

A decision tree can be used to visually represent decisions and decision making. Both classification trees and regression tress can be drawn, and such trees are examples of supervised learning. Typically, the decision tree is drawn upside down with its root at the top, going from observations about an item/object to conclusions about the item's classification/value. Examples of uses of decision trees include classifying types of birds based on factors such as size, colour, markings, country, or using features of a house and neighbourhood to determine possible selling price.

**SVM (Support Vector Machine)**

These are supervised learning models that can be used for both regression and classification tasks. The method aims to find the maximum distance between data points of different classes. To find this maximum distance hyperplanes (decision boundaries that help classify the data points) are used. Data points falling on either side of the hyperplane can be attributed to different classes i.e., the hyperplane is the decision boundary. Support vectors are data points that are closer to the hyperplane and so influence the position and orientation of the hyperplane. Effectively, after giving an SVM model a set of labelled data for each category, the model is them able to categorize new data. One example of its use it to classify the topic of a passage of text.

**Naïve Bayes**

These are supervised learning models used for prediction. It is a method often used to classify data by using the Bayes theorem of probability to predict the class of unknown data. Examples of its uses include identifying spam, and classifying items (e.g., groups of customers, fruit etc.)

**K-Means**

These are unsupervised learning models and are clustering algorithms, where the data is grouped into different classes without referring to known, or labelled, outcomes. It groups similar data points together, in order to discover underlying patterns. K-means looks for a fixed number (k) of clusters (a collection of data points grouped together based on certain similarities) in a dataset.