|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Machine Learning Algorithm** | **Is it Supervised/ Unsupervised/ Reinforcement learning?** | **What does the algorithm do?** | **In which situations will it be most useful?** | **Can you find any examples of where this algorithm has been used?** |
| Linear Regression | Supervised | Finds the relationship between two variables on a -1 to 1 scale (with a scalar var input). + meaning as one increase the other is likely to aswell. – meaning as one increase the other is likely to decrease. Values closer to zero suggest there is no relationship. (-+1-0.5 strong, +-0.4-0.2 med below that is weak) | Predicting / Forecasting | Sales – where elasticity of demand is higher, what is the expected impact on sales from increasing/decreasing price? Lower price should increases sales volume and vice versa (up to a certain point) |
| Logistic Regression | Supervised | Similar to linear but works for categorical data (binary, nominal or ordinal) with multiple inputs. | Prediction / Forecasting where there is only 1 of 2 outcomes | Voting – Which demographics are indicative of voting A or B to either predict the outcome of an election based on anticipated volume of voters or to target/prioritise resources to swing voters to shift them towards voting a particular way.  Employee engagement/ staff retention: Which variables are associated with the decision to leave a business? Which have the most impact? E.g. The experience of a direct line manager can be more impactful than that of senior leaders. How low a score on manager rating makes an employee a flight risk? |
| Decision Tree | Supervised | Modelling the decisions / branches and assigning probability to calculate the chances of various outcomes. | Prediction / Forecasting | Behavioural analysis: What’s the likelihood of an event e.g. Will a customer buy a product, Is it known to them, have they seen advertising, did they receive a coupon as part of targeted advertising, where is the item positioned, what time of the month is it. |
| SVM (Support Vector Machine) | Supervised but can be Unsupervised | A combination of regression and clustering useful with smaller samples that have a range of descriptive variables. With a supervised approach the data can be labelled to give it some training points but can also be done unsupervised. Samples need to have some sort of distinction between them otherwise it will struggle visually to separate into something useful. | Classification | Images – grouping images based on characteristics Sales – grouping potential customers in suitable packages. E.g. mortgage applications can be grouped based on the user demographics etc which will cluster them based on risk etc determining the amount to borrow, for how long and at what rate. |
| Naive Bayes | Supervised | Classification of new sample based on previous examples by calculating the probability of traits amongst the initial set and testing new sample against it. | Classification | Spam filters – Emails flagged as spam/phishing/genuine are used to establish the probability of new email being spam. E.g. Spam tends to have many hyperlinks, mismatched sender vs sent from name, sent at unusual times, reference certain words. If a new email meets the key traits associated with spam, it’ll get flagged. |
| KNN (K-Nearest Neighbours) | Supervised | Classification of new sample based on proximity to a specified number of neighbours (K) from a known set. | Classification | Financial risk modelling – using a combination of home owner status, age, marital status etc, is an applicant more or less likely to default on their borrowing based on previous experience of people in a similar situation. |
| K-Means |  |  |  |  |
| Random Forest |  |  |  |  |