**MACHINE LEARNING**

1. Which machine learning algorithm(s) can be used for my specific problem?

This is a fundamental question we should always ask and find the answer to before using any machine learning algorithm.

This blog post will share what factors should be considered and the questions that need to be answered in order to select the most appropriate machine learning algorithm(s) to solve the problem at hand.

Machine learning algorithms are divided into three types: **supervised**, **unsupervised** and **reinforcement learning**.

***Question one: what type of machine learning algorithm should I select?***

1.      **Supervised learning**: we have labelled datasets, and the goal is to predict the future outcomes.

2.**Unsupervised** **learning**: data points have no labels associated with them, and the goal is to organize the data in some way or discover the intrinsic patterns that underlie the data.

3.      **Reinforcement learning**: the algorithm gets to choose an action in response to each data point with the goal of achieving the highest reward based on how good the decision of the algorithm was.

After deciding if the problem is supervised, unsupervised or reinforcement learning we need to look at the available algorithms under each of these categories. Here are some examples of supervised and unsupervised algorithms:

**Supervised:**

1. Classification for predicting a categorical variable:

* Linear SVM
* Naive Bayes
* Decision Tree
* Logistic Regression
* Kernel SVM
* Random Forest
* Neural network
* Gradient Boosting Tree

2. Regression for predicting a continuous variable:

* Linear Regression
* Random Forest
* Neural network
* Gradient Boosting Tree
* Decision Tree

**Unsupervised:**

1. Clustering for grouping data in a way that observations in one group or cluster are more similar (according to some criteria) than those in other groups:

* k-means
* k-modes
* Gaussian Mixture Model
* DBSCAN

2. Dimension reduction for reducing the number of variables under consideration, specifically, the redundant or irrelevant features:

* Principal Component Analysis
* Singular Value Decomposition
* Latent Dirichlet Analysis
* Linear Discriminant Analysis (LDA)
* Autoencoders

There is a long list of algorithms available under each type of machine learning algorithms. It is worth noting that there is not one algorithm that works best for every problem since there are many factors at play.

***Question two: which algorithm works best for my specific problem?***

To determine other factors that should be considered when selecting a machine learning algorithm, the following questions can help narrow your options into a smaller group:

**What is the acceptable accuracy level?**

Typically, there is a trade-off between accuracy and training time. If for your case approximation is good enough, you should select algorithms with approximate methods. This helps to save processing time.

**What is the available training time?**

If you have enough time to spend on training and building the model, then you can move to algorithms that result in higher accuracy.

**What is the size of data?**

The size of data also affects the accuracy and training time of the algorithms.

**What are the assumptions made by the algorithms?**

Each algorithm comes with some assumptions. You should ensure you meet the algorithm’s requirements; otherwise, it will reduce the accuracy of your model. These assumptions may be about the number of observations, the relationship between features, the limit on the number of categories, whether features are linear or nonlinear and whether feature values are discrete or continuous, to name a few.

**What are the number of parameters?**

The more parameters you have for an algorithm, the more time you should spend configuring them and finding the best combination that leads to the highest accuracy.

**What are the number of features?**

If the number of features is large compared to the number of data points, then only certain algorithms can be used.

**What are the memory requirements?**

Memory usage of an algorithm is the other factor we should check if we have limited memory available.

**What is the result interpretation/explain-ability of the algorithm?**

Some algorithms are more complex than others, therefore it is more difficult to interpret their result. Depending on if you are looking for easy or difficult result interpretation, some algorithms should be selected over others.

**Final note:** We can use the criteria above to limit the number of algorithms we use to resolve our problem. However, usually more than one algorithm can solve a given problem, and one algorithm may be a better fit than others. So, in cases where it is not possible to know the best algorithm in advance, we can try one of the other candidates, and if the results are not acceptable, we can try the other ones.