How ML algorithms works?

- Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y). Y = f(X)
- This is a general learning task where we would like to make predictions in the future (Y) given new examples of input variables (X). We don’t know what the function (f) looks like or it's forming.
- If we did, we would use it directly and we would not need to learn it from data using machine learning algorithms. It is harder than you think. There is also error (e) that is independent of the input data (X).

  Y = f(X) + e
- This error might be error such as not having enough attributes to sufficiently characterize
  the best mapping from X to Y.

How ML algorithms works?

- This error is called irreducible error because no matter how good we get at estimating the target function (f), we cannot reduce this error.

Techniques For Learning a Function

- Machine learning algorithms are techniques for estimating the target function (f) to predict the output variable (Y) given input variables (X). Different representations make different assumptions about the form of the function being learned, such as whether it is linear or nonlinear.
- Different machine learning algorithms make different assumptions about the shape and structure of the function and how best to optimize a representation to approximate it.
- This is why it is so important to try a suite of different algorithms on a machine learning problem because we cannot know beforehand which approach will be best at estimating the structure of the underlying function we are trying to approximate.

Parametric and Nonparametric ML Algorithms

Parametric Machine Learning Algorithms

- A learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples). No matter how much data you throw at a parametric model, it won't change its mind about how many parameters it needs. as we have a finite number of parameters.
- Linear models such as linear regression, logistic regression, and linear Support Vector Machines are typical examples of a parametric "learners;" here, we have a fixed size of parameters (the weight coefficient.)
- some more examples Naive Bayes, Perceptron, simple Neural network without weight decay.

Parametric and Nonparametric ML Algorithms

Advantages of Parametric Machine Learning Algorithms:

- **Simpler:** These methods are easier to understand and interpret results.
- **Speed:** Parametric models are very fast to learn from data.
- **Less Data:** They do not require as much training data and can work well even if the fit the data is not perfect.

Disadvantages of Parametric Machine Learning Algorithms:

- **Constrained:** By choosing a functional form these methods are highly constrained to the specified form.
- **Limited Complexity:** The methods are more suited to simpler problems.
- **Poor Fit:** In practice, the methods are unlikely to match the underlying mapping function

Parametric and Nonparametric ML Algorithms

Nonparametric Machine Learning Algorithms

- Nonparametric methods are good when you have a lot of data and no prior knowledge, and when you don’t want to worry too much about choosing just the right features.
- In nonparametric models, the number of parameters is (potentially) infinite and the complexity of the model grows with the number of training data.
- In contrast, K-nearest neighbour, decision trees, or RBF kernel SVMs are considered as non-parametric learning algorithms since the number of parameters grows with the size of the training set.
- More examples Decision Trees like CART and C4.5, Naive Bayes, Support Vector Machines, Neural Networks with weight decay.

Parametric and Nonparametric ML Algorithms

Nonparametric Machine Learning Algorithms

Advantages of Nonparametric Machine Learning Algorithms

- **Flexibility:** Capable of fitting a large number of functional forms.
- **Power:** No assumptions (or weak assumptions) about the underlying function.
- **Performance:** Can result in higher performance models for prediction.

Disadvantages of Nonparametric Machine Learning Algorithms:

- **More data:** Require a lot more training data to estimate the mapping function.
- **Slower:** A lot slower to train as they often have far more parameters to train.
- **Overfitting:** More of a risk to overfit the training data and it is harder to explain why specific predictions are made.