Parametric vs Non-Parametric Model:

<https://machinelearningmastery.com/parametric-and-nonparametric-machine-learning-algorithms/>

Parametric Model: *A learning model that summarizes data with a set of parameters of fixed size (independent of the number of training examples) is called a parametric model. No matter how much data you throw at a parametric model, it won’t change its mind about how many parameters it needs.*

Some more examples of parametric machine learning algorithms include:

* Logistic Regression
* Linear Discriminant Analysis
* Perceptron
* Naive Bayes
* Simple Neural Networks

Benefits 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 to the data is not perfect.

Limitations 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.

Non Parametric Model: Algorithms that do not make strong assumptions about the form of the mapping function are called nonparametric machine learning algorithms. By not making assumptions, they are free to learn any functional form from the training data.

*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.*

Some more examples of popular nonparametric machine learning algorithms are:

* k-Nearest Neighbors
* Decision Trees like CART and C4.5
* Support Vector Machines

Benefits 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.

Limitations 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.
  1. Machine learning?

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Major Machine learning algorithms:

Regression: Predicting continuous values

* + - Linear Regression
    - Polynomial Regression
    - Exponential Regression
    - Logistic Regression
    - Logarithmic Regression

Classifications: predicting set of category

* + - K-Nearest Neighbours
    - Decision trees
    - Random forests
    - Support Vector Machines
    - Naïve Bayes

Clustering: Hidden clusters in the data

* + - K-means
    - DBSCAN
    - Mean Shift
    - Hierarchical

Association: Associating co-occurring items or events

* + - Apriori

Anomaly Detection: Discovering abnormal activities and unusual cases like fraud detection

Sequence Pattern Mining: Predicting next data events between data examples in a sequence

Dimensionality Reduction: Reducing the size of data to extract only useful features from a dataset.

Recommendation System:

* 1. What is Underfitting, Overfitting and balanced?

**Underfitting:** Model performs poorly on training set. Underfit learners tend to have low variance but high bias. The model simply does not campture the relationship of the training data, leading to inaccurate predictions of the training data.

**Causes:**

* + - Trying to create a linear model with non linear data.
    - Having too little data to build an accurate model
    - Model is too simple, has too few features

**Remedies:**

* + - Add more features to the model.
    - Engineer additional features which makes sense.

**Overfitting:** Model performs too well on training set but fails on the test set. An overfit model shows low bias and high variance. The model is excessively complicated likely due to redundant features.

**Causes:**

* + - Model has captured noise
    - Trained on less data

**Remedies:**

* + - Regularization.
    - K-fold Cross Validation.
    - Train with more data.
    - Remove features.
    - Ensembles
  1. Unsupervised Learning:

Basic type of unsupervised learning is dimension reduction methods, such as PCA, t-SNE, while PCA is generally used in data preprocessing, and t-SNE usually used in data visualization.

A more advanced branch is clustering, which explores the hidden patterns in data and then makes predictions on them; examples include K-mean clustering, Gaussian mixture models, hidden Markov models, and others.

Examples of unsupervised learning algorithms:

* Dimensionality Reduction
* Densith Estimation
* Market Basket analysis
* Generative adversarial networks(GANs)
* Clustering