Modern Machine Learning Algorithms

The Big 3

REMEMBER
- Practice, practice, practice.
- Others algs (most) = Adaptaions so master this ones (the fundamentals).
- Algorithms = commodities.

NO FREE LUNCH
no algorithm works best for every problem. There are many factors at play to already state which algorithm is the best. This is why you should try many different algirithms for your problem. Of course, the algorithms you try should be appropiated for your problem, and that's when picking the right machine learning task comes in.

CLUSTERING
- **K-Means**
- **Afinity Propagation**
- **Hierarchical - Agglomerative**
- **DBSCAN**

Unsupervised learning task for finding natural groupings of observations based on the inherent structure within the dataset.

- makes clusters based on geometric distance between points.
- Makes clusters based on graph distances between points.
- Clusters in groups until it leave you with a hierarchy of clusters.
- DBSCAN makes clusters for dense regions of points. HDBSCAN allows varying density clusters.

REGRESSION
- **(Regularized) Linear Regression**
- **Regression Trees (Ensembles)**
- **Deep Learning**
- **Nearest Neighbors**

Is the supervised learning task for modeling and predicting continuos, numeric variables.

- fit a straight hyperplane or line of the relationship of the variables
- Branching structure that allows it to fit non-linear relationships.
- Multi-layer networks that can learn extremely complex patterns.
- instance-based algorithms = they save each training observation. They then make predictions for new observations by searching for the most similar training observations and pooling their values.

CLASSIFICATION
Like reression BUT for caregorical data.
- **(Regularized) logistic regression**
- **Classification Trees (Ensembles)**
- **Deep Learning**
-
- **2.4 - Support Vector Machines**
- **Naive Bayes**

- classification counterpart of linear regression.
- Counterpart of regression trees.
- Easily adapted to classification problems. Common use.
- max(distance) between the closest members of separated classes.
- probability table that gets updated through your training data.