**Supervised Learning:** Dataset is split into training and test data with known labels. Model undergoes “training” until a certain desired accuracy is reached.

Goal: Reach a certain accuracy.

Problems: Classification, Regression

Algorithms: Logistic Regression, etc

**Unsupervised Learning:** Input data is not labeled, no known desired result.

Goal: To find trends and patterns in relationships between data.

Problems: Clustering, Dimensionality Reduction, Association Rule Learning.

Algorithms: Apriori algorithm and K-Means.

**Classification:** Taking an input and assigning a label to it. Usually, output is discrete (categorical), of a “yes or no” nature

**Regression:** Used to find labels of a numerical (continuous) nature like “How many?”

**Ensembling:** Combining the predictions of multiple machine learning models that are individually weak to produce more accurate models

**Overfitting:** Modeling error that occurs when a machine learning model performs extremely well on training data, but poorly on unseen, or test data. Often a result of

* Excessively complicated (many features) model
* Low bias but high variance

**Underfitting:** Modeling error that occurs when a machine learning model does not perform well in tests; does not fit the data well enough. Often a result of

* Not enough features or missing key features to determine a trend
* High bias but low variance

**Noise:** Irrelevant information or randomness in a data set