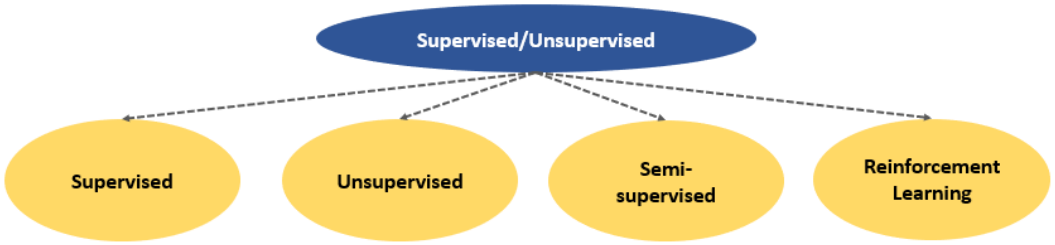
Supervised vs Unsupervised

Tuesday, December 3, 2019 7:21 PM

Note Written by: Zulfadli Zainal

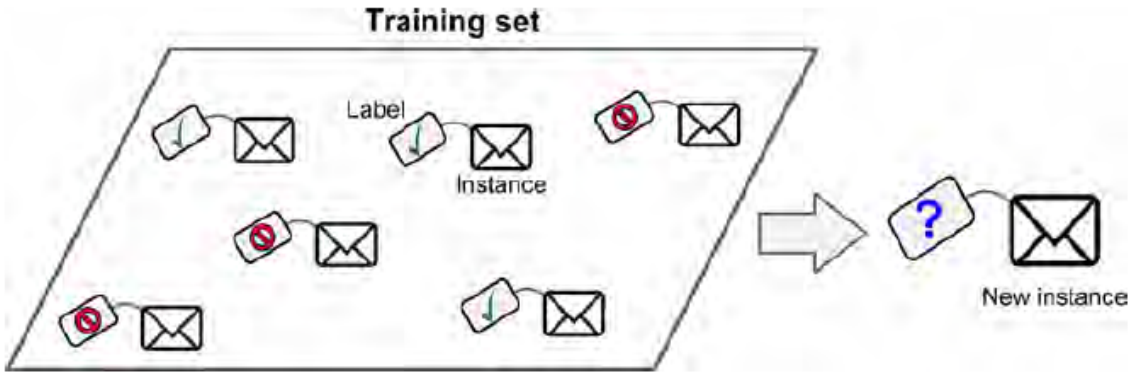
Based on the level of human supervision, we can categorize this type with 4 major category.



Supervised Learning

In supervised learning, the training data includes desired solution (labels).

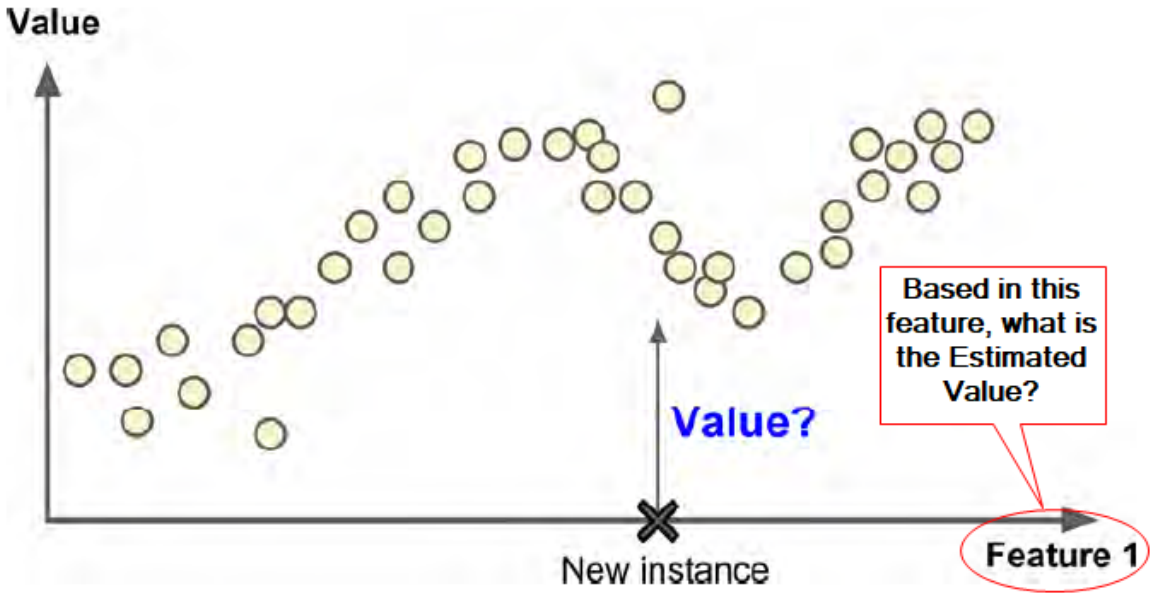
The training data is labelled with desired output.



Typical supervised learning task is Classifications & Regression.

Classification: Train many sample based on Class.
Regression: Train many sample based on Features and Labels.

Sample Regression:

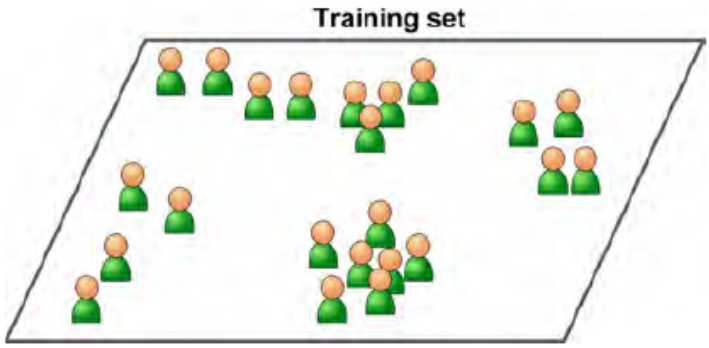


Important Supervised Learning Algorithm:

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks

Unsupervised Learning

The training data is not labeled. We are trying to teach the machine without any teacher!



Important Unsupervised Learning Algorithms

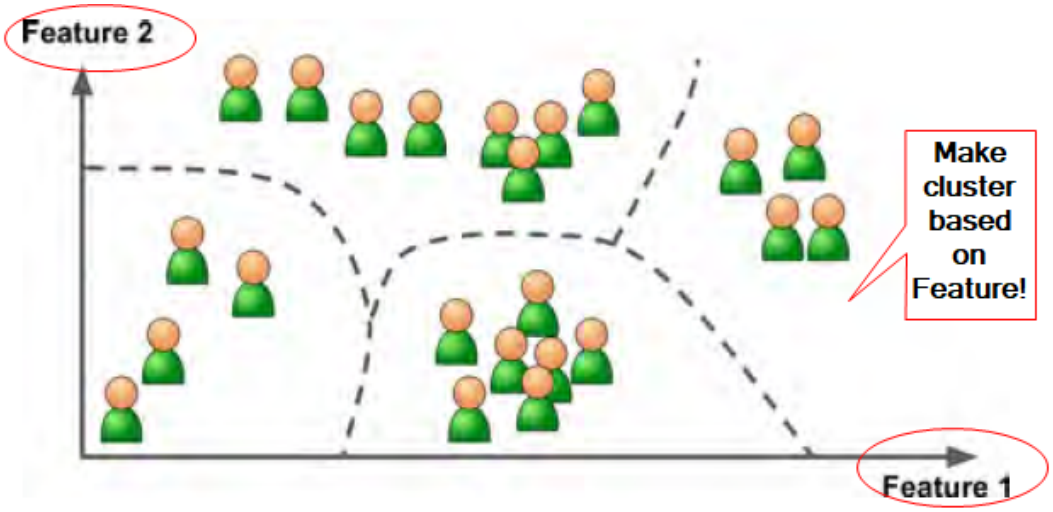
• Clustering
—k-Means
—Hierarchical Cluster Analysis (HCA)
—Expectation Maximization

• Visualization and dimensionality reduction
—Principal Component Analysis (PCA)
—Kernel PCA
—Locally-Linear Embedding (LLE)
—t-distributed Stochastic Neighbor Embedding (t-SNE)

• Association rule learning
—Apriori
—Eclat

For example (Clustering):

In case of **clustering**: you don't have to tell the machine how to group the data, but the machine will analyze the data and cluster them based on group (and also subgroup).

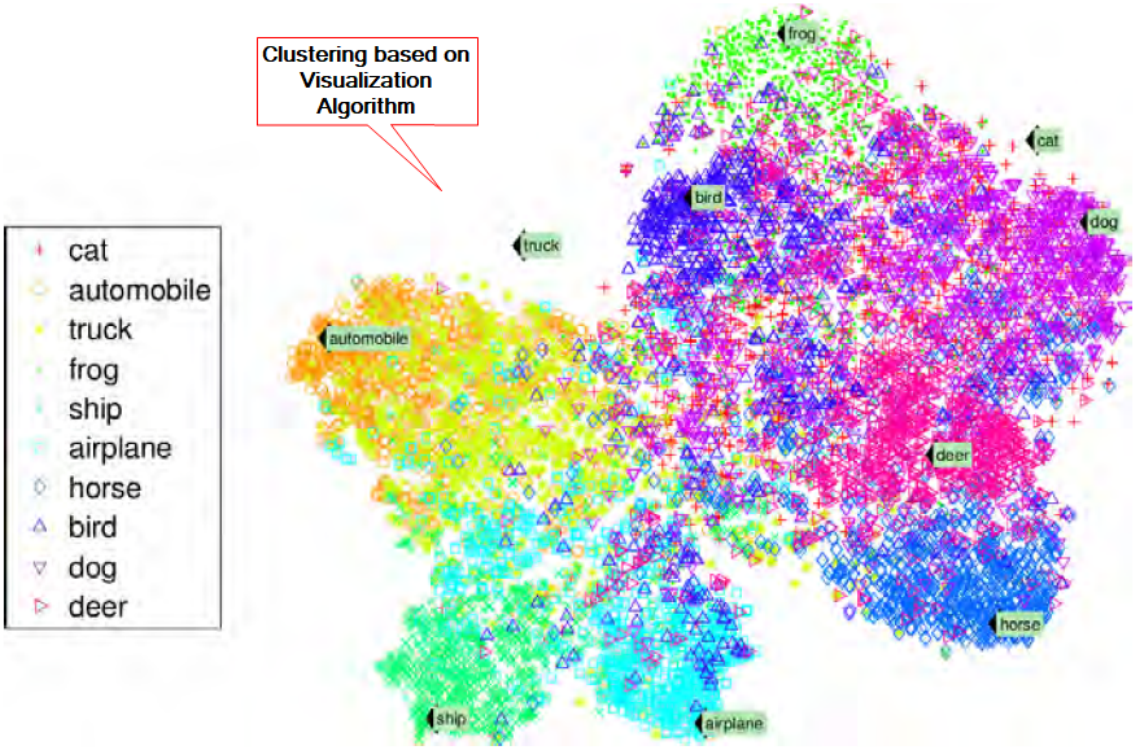


For example (Visualization Algorithm):

Visualization Algorithm is good examples of unsupervised learning algorithms.

You feed them a lot of complex and unlabeled data, and they output a 2D or 3D representation of your data that can easily be plotted. These algorithms try to preserve as much structure as they can.

Why important: So its easy for us to understand the data!



For example (Dimension Reduction Algorithm):

The goal is to simplify the data without losing too much information. Eg: dimensionality reduction algorithm will merge several very related feature into one feature.

It will run much faster, the data will take up less disk and memory space, and in some cases it may also perform better.

For example (Anomaly Detection Algorithm):

Example: If the data not following certain trend.

For example (Association Rule Learning Algorithm):

The goal is to dig into large amounts of data and discover interesting relations between attributes.

Semi-Supervised Learning

Some algorithms can deal with partially labeled training data -> This is Semi Supervised!

Eg: Google Photos (Unsupervised: when detecting similar faces, Supervised: User label who is in the picture)

Reinforcement Learning

Reinforcement Learning is a very different beast.

Using the concept of 'Agent' that observe the environment -> Reward (if =ve) and Penalties (if -ve). From this rewarding system, the machine will create a policy where it can get the most reward over time.