INTRODUCTION TO MACHINE LEARNING

LET'S HAVE A LOOK AT SOME USE CASES....

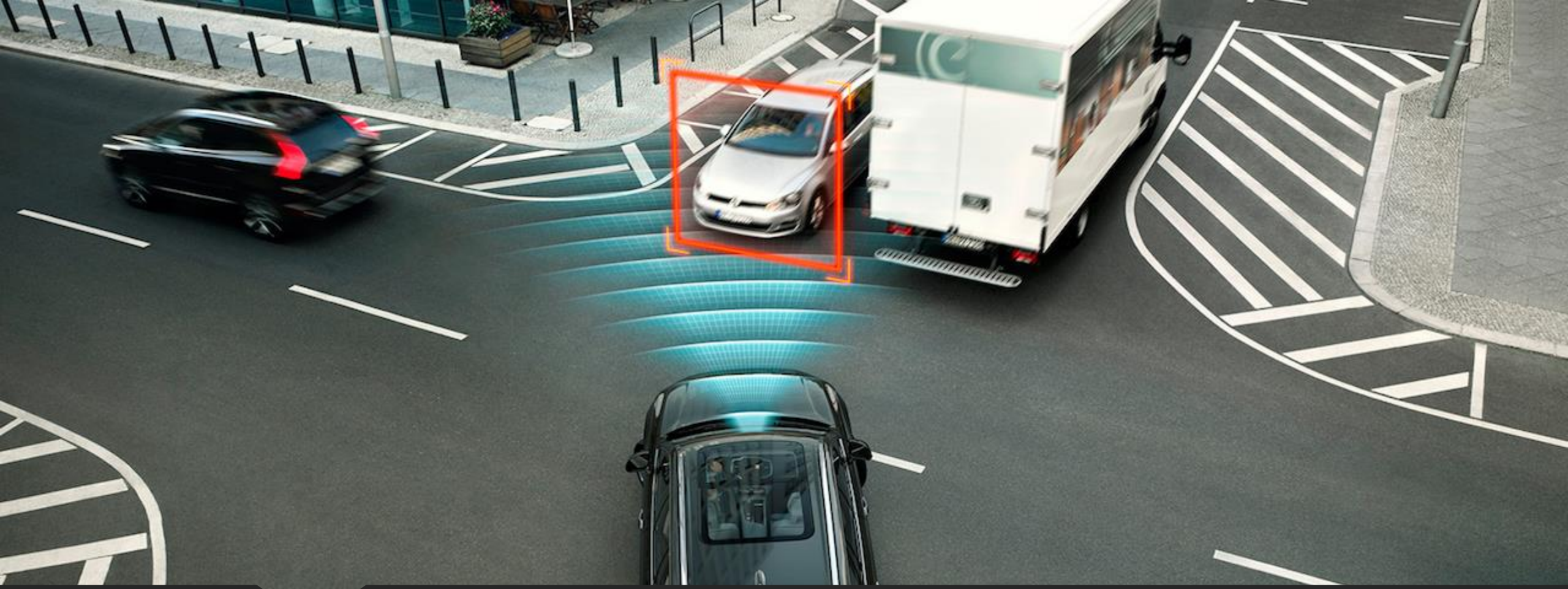
- Recommendation Systems
- Face Recognition
- Self Driving Cars
- Voice Assistants

YOUTUBE RECOMMENDATION SYSTEM

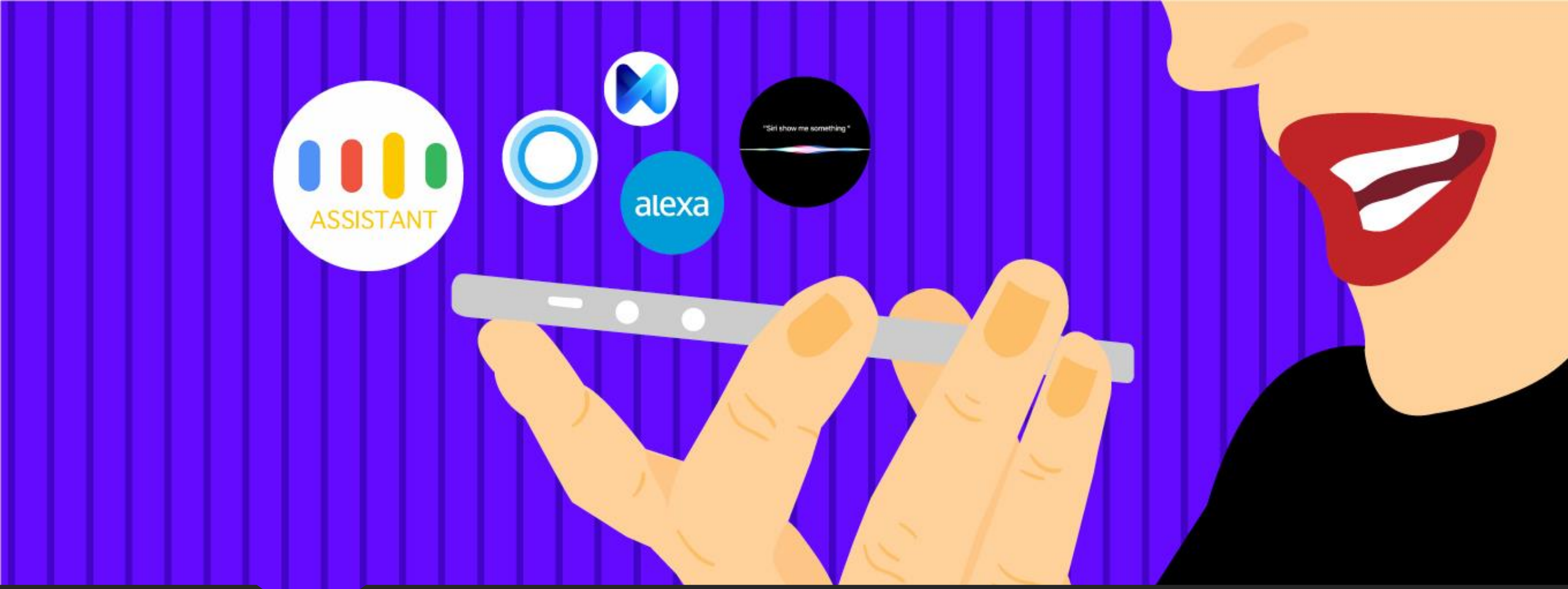
INSTORE RECOMMENDATION SYSTEMS

FACEBOOK IMAGE TAGGING

IPHONE X FACE RECOGNITION ID LOCK



SELF DRIVING CARS



VOICE ASSISTANTS (SIRI, GOOGLE, ETC.)

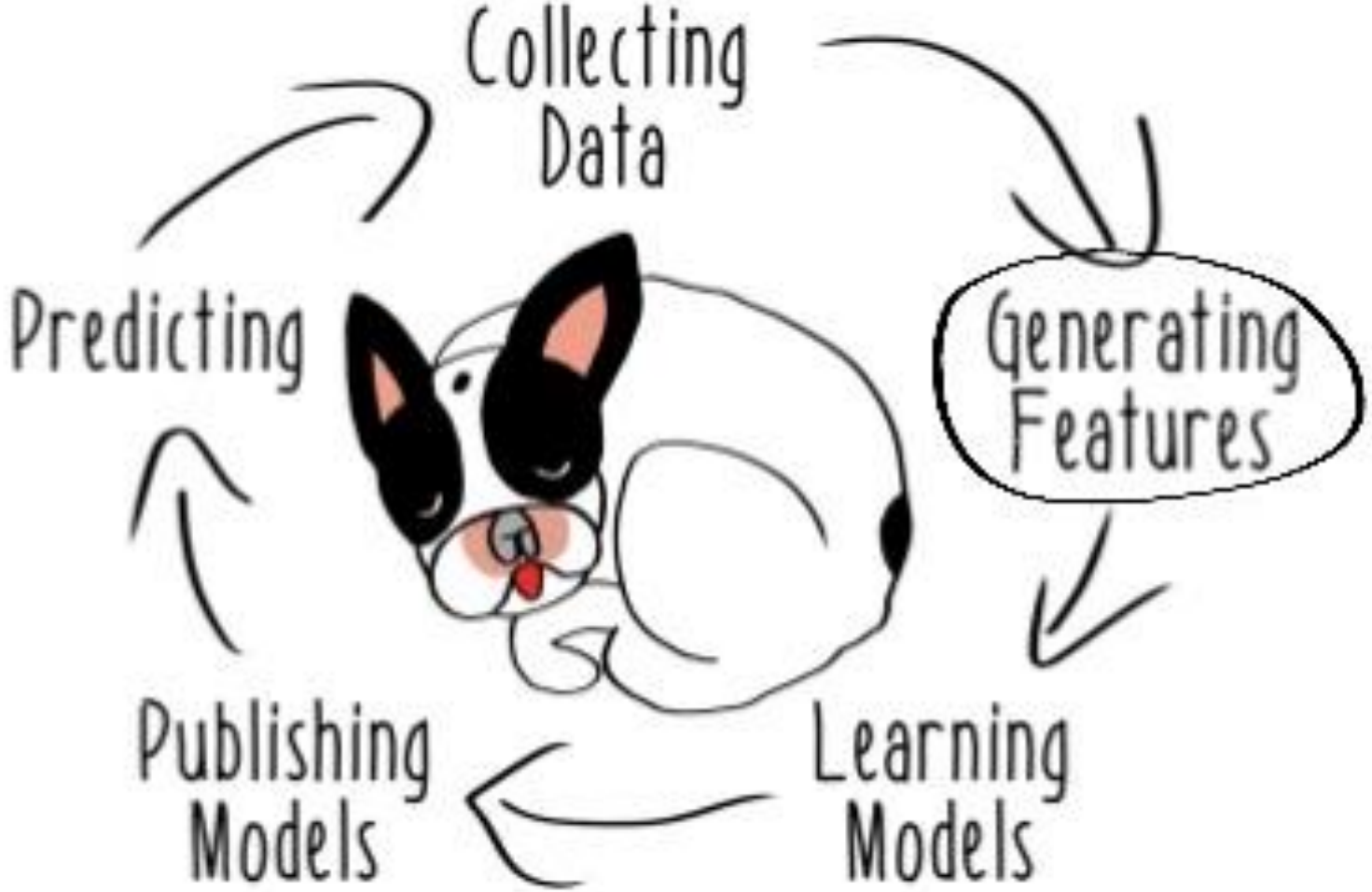
BUT WHAT IS MACHINE LEARNING?

- Computational learning using algorithms to learn from and make **predictions** on data.
- Iteratively learn from data.
- Find hidden insights.
- Types of datasets:

  'Train the model on' --> TRAINING SET

  'Test the model on' --> TEST SET

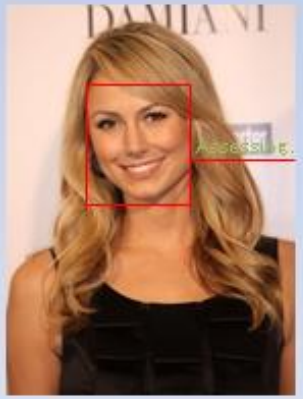
Machine Learning Systems



MACHINE LEARNING METHODS…

- Supervised methods
- Unsupervised methods

Supervised Learning



Unsupervised Learning

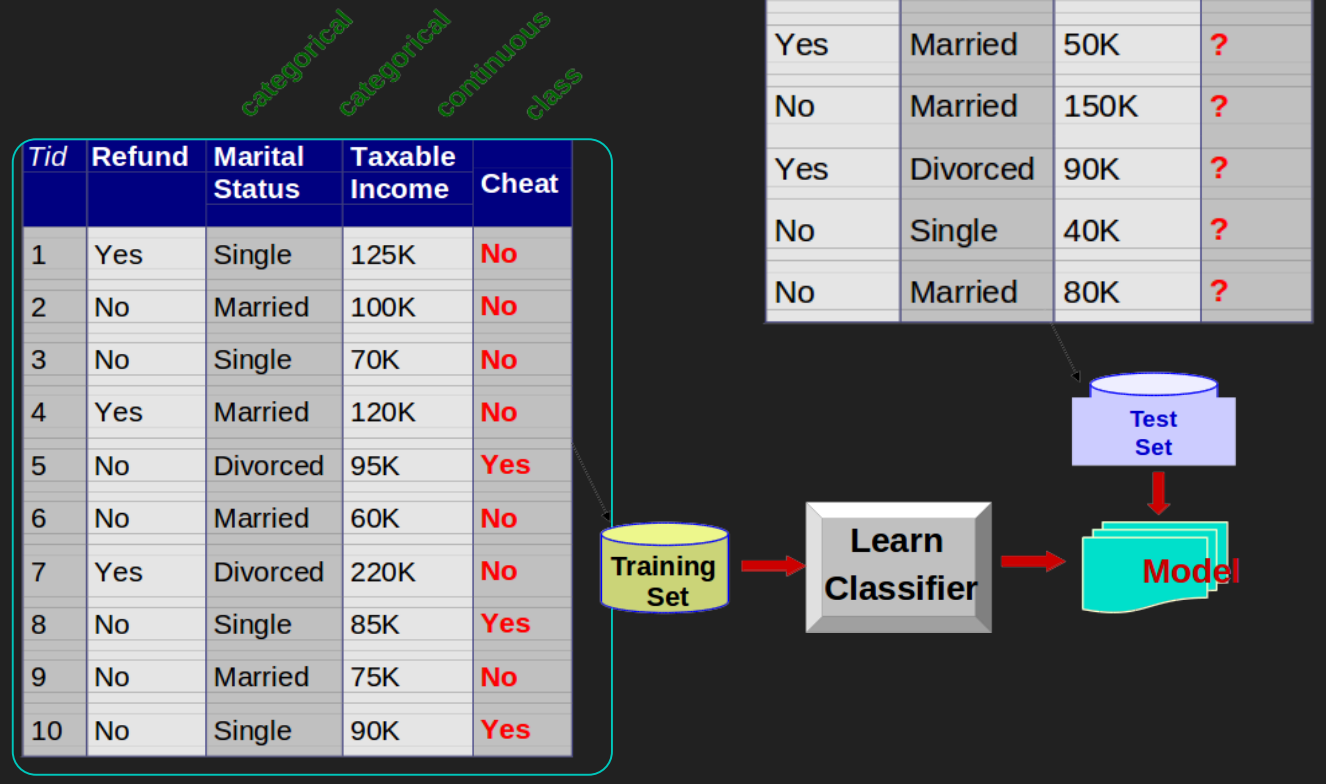
Kontagent

SUPERVISED METHODS

- Data comes with attributes that are to be predicted.
- Task is well-defined and structured.
- Classification and Regression are most common supervised methods.

CLASSIFICATION

- Samples belong to two or more classes.
- Learn from already labeled data how to predict the class of unlabeled data.
- Discrete form of supervised learning.
- Example- Cats and Dogs, Sentiment Analysis



REGRESSION

- Predict a value of a given continuous valued variable based on the values of other variables.
- Greatly studied in statistics, neural network fields.
- Examples:

  Predicting sales amounts of new product based on advertising expenditure.

  Predicting wind velocities as a function of temperature, humidity, air pressure, etc.

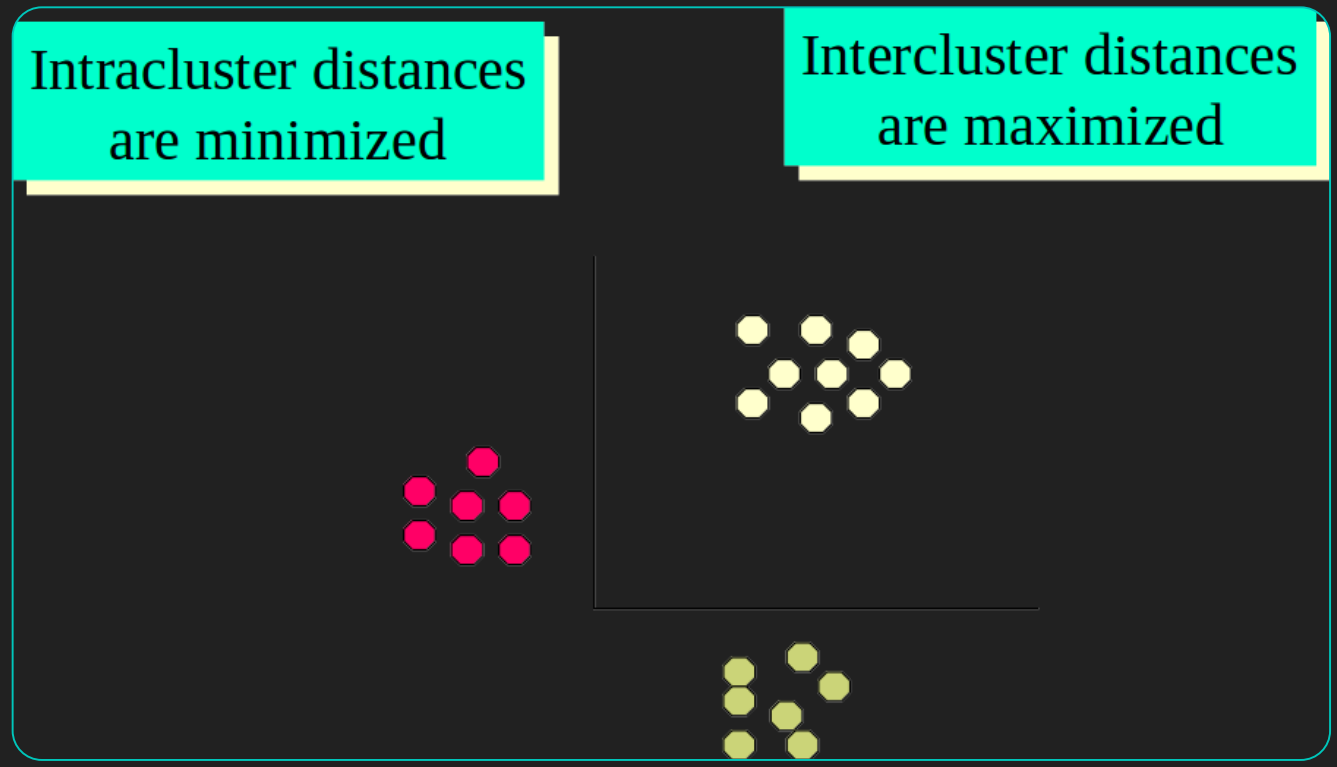
  Time series prediction of stock market indices.s

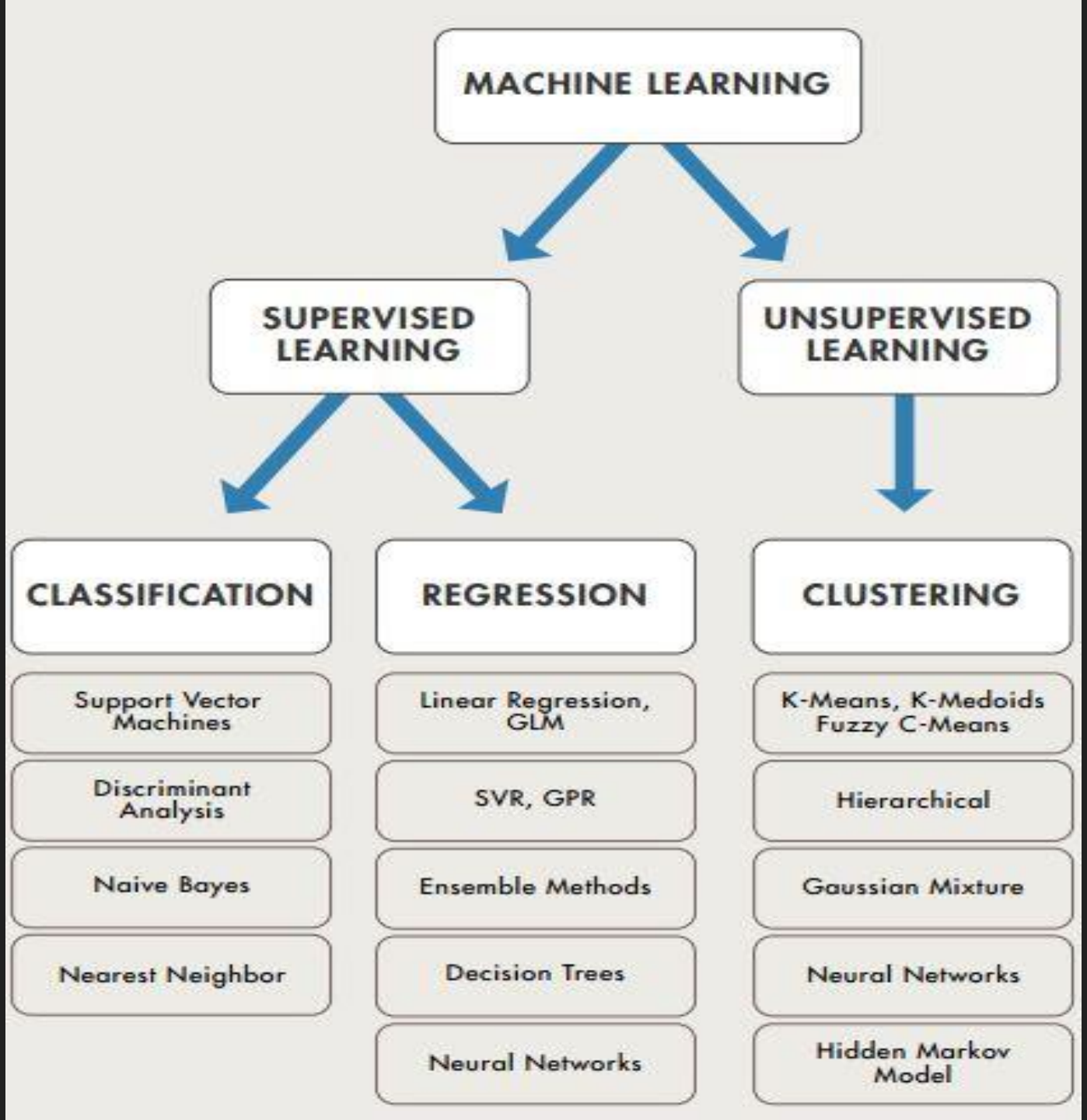
UNSUPERVISED METHODS

- Training data consists of a set of input without any corresponding target values.
- Clustering is a common unsupervised method.

CLUSTERING

- Segregate groups with similar traits and assign them into clusters.
- Example- Clusters of customers with similar shopping habits.




MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
CLASSIFICATION
REGRESSION
CLUSTERING
Support Vector Machines
Discriminant Analysis
Naive Bayes
Nearest Neighbor
Linear Regression, GLM
SVR, GPR
Ensemble Methods
Decision Trees
Neural Networks
K-Means, K-Medoids Fuzzy C-Means
Hierarchical
Gaussian Mixture
Neural Networks
Hidden Markov Model

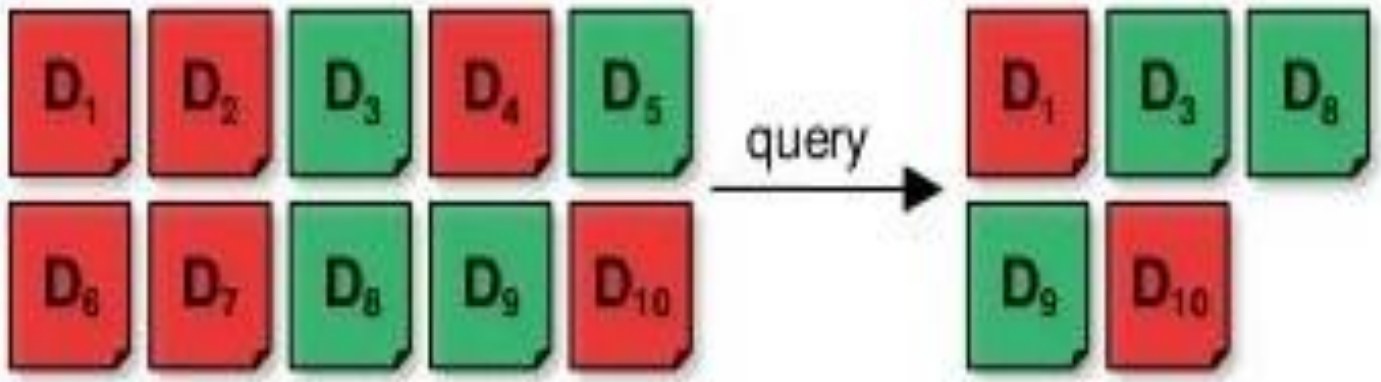
EVALUATION OF MODEL

Commonly used metrics:

- Accuracy: Proportion of the total number of predictions that are correct.
- Precision: Fraction of relevant instances among the retrieved instances.
- Recall: Fraction of relevant instances that have been retrieved over the total amount of relevant instances.

$$\text{precision} = \frac{|\{\text{relevant documents}\}| \cap |\{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\}| \cap |\{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$



$$\text{precision} = \frac{3}{5} = 0.6$$

$$\text{recall} = \frac{3}{4} = 0.75$$

Resources

- Video Tutorial- https://www.coursera.org/learn/machine-learning
- Scikit Learn (Library for Machine Learning in Python) Tutorial- https://github.com/amueller/scipy-2016-sklearn
- Blog posts- https://medium.com/machine-learning-for-humans/how-to-learn-machine-learning-24d53bb64aa1

THANK YOU!