Plan:

1. Define supervised and unsupervised ML
2. Explain difference between classification and regression
3. Conceptually explain various classic ML models

Machine Learning: Classic Models

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

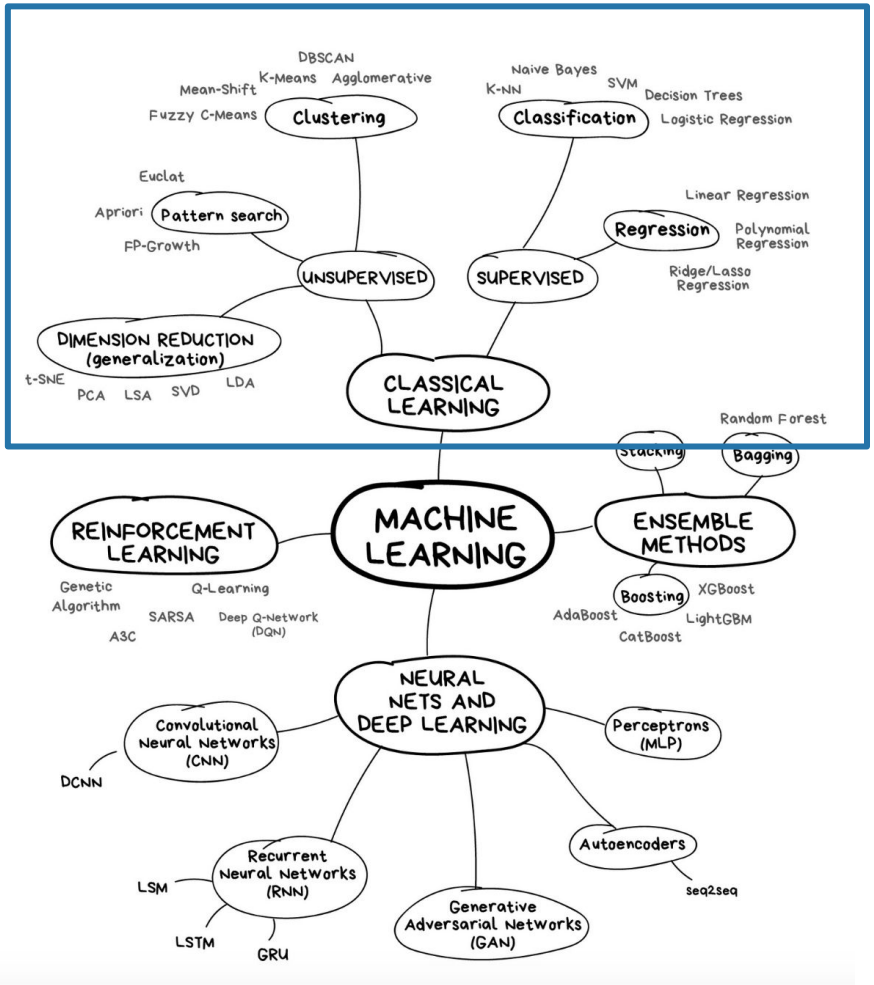
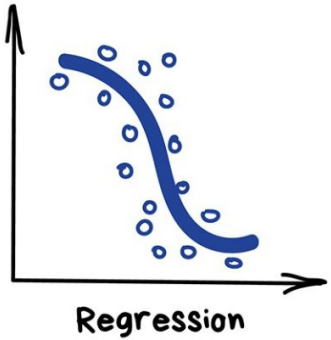


Image source: https://vas3k.com/blog/machine_learning/

Regression

"Draw a line through these dots. Yep, that's the machine learning"



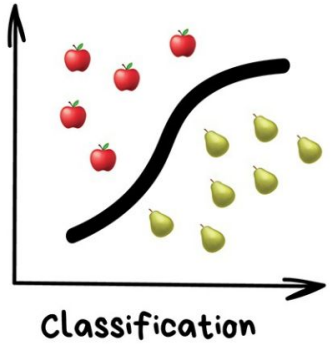
Today this is used for:

- Stock price forecasts
- Demand and sales volume analysis
- Medical diagnosis
- Any number-time correlations

Popular algorithms are Linear and Polynomial regressions.

Classification

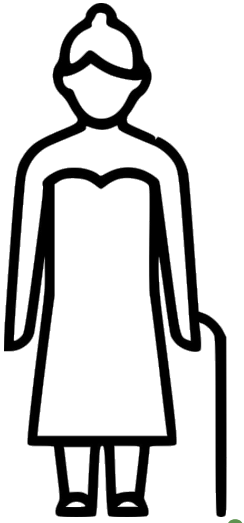
"Splits objects based at one of the attributes known beforehand. Separate socks by based on color, documents based on language, music by genre"



Today used for:
– Spam filtering
– Language detection
– A search of similar documents
– Sentiment analysis
– Recognition of handwritten characters and numbers
– Fraud detection

Popular algorithms: Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine

Image source: https://vas3k.com/blog/machine_learning/



Regression:

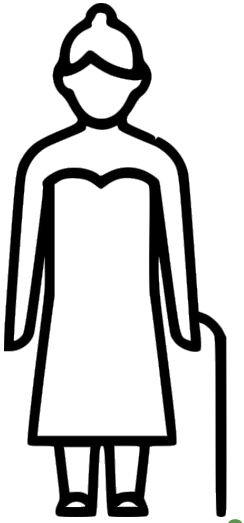
predicting continuous variables
(i.e. Age)

continuous variable prediction

Classification:

predicting categorical variables
(i.e. education level)

categorical variable prediction



Regression:

predicting continuous variables
(i.e. Age)

continuous variable prediction

Classification:

predicting categorical variables
(i.e. education level)

categorical variable prediction

Supervised Learning

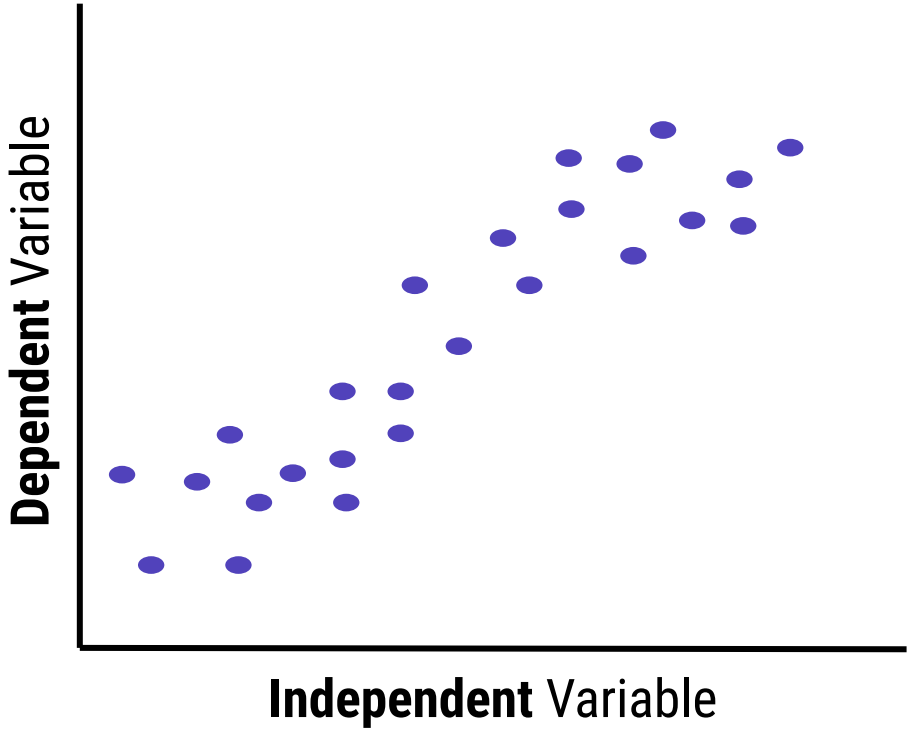
regression

continuous variable prediction

Supervised Learning

regression

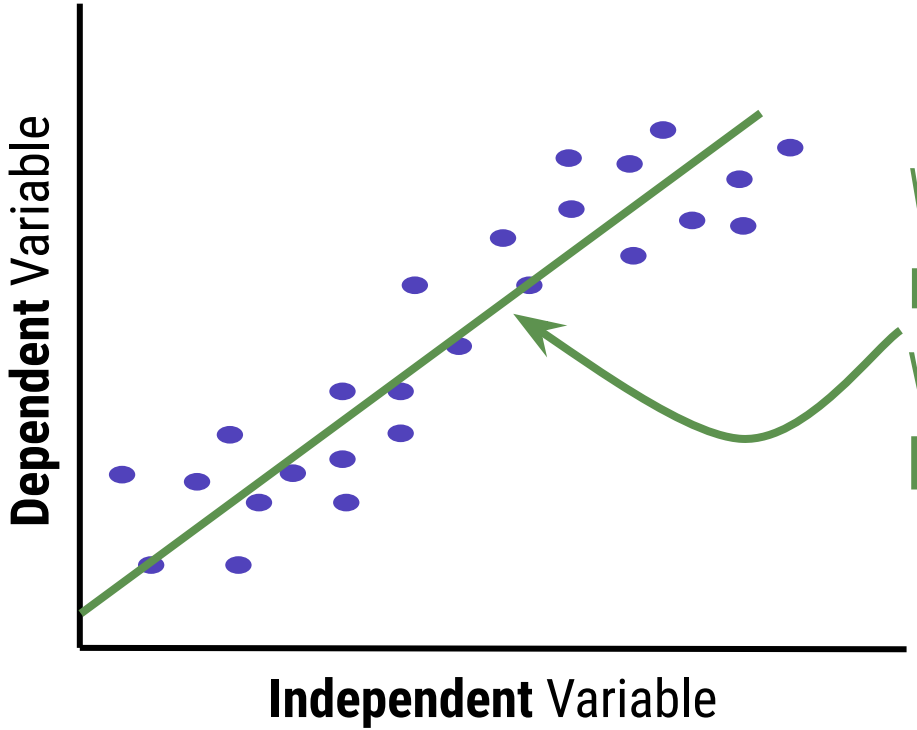
continuous variable prediction



Supervised Learning

regression

continuous variable prediction

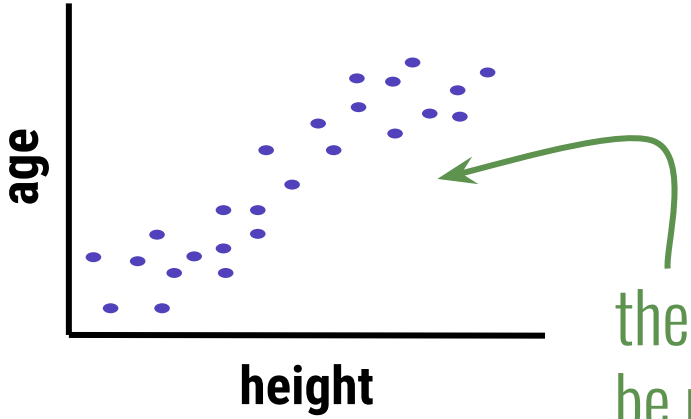


We'll use the linear relationship between variables to generate a **predictive model**

Supervised Learning

regression

continuous variable prediction

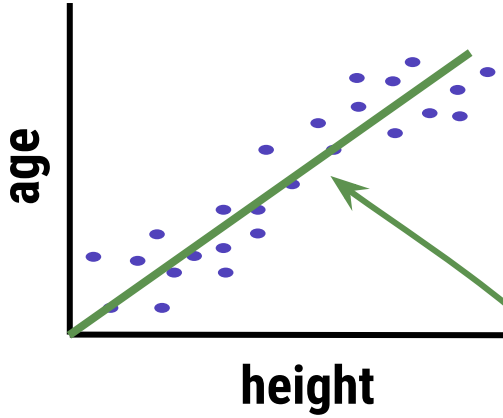


the training data will be used to build the predictive model

Supervised Learning

regression

continuous variable prediction

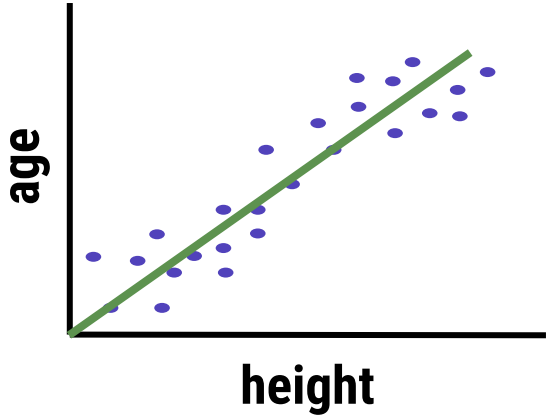


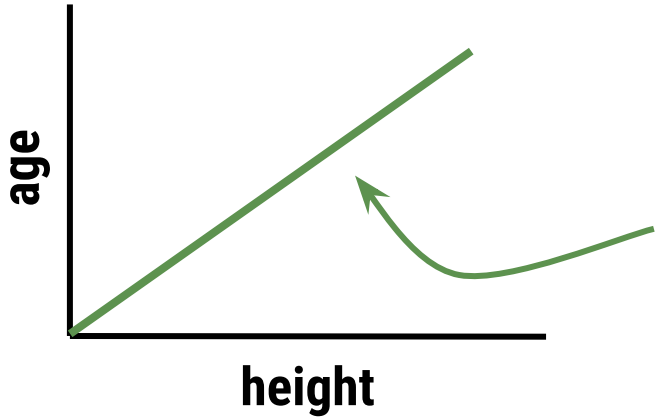
use linear regression to model the relationship

Supervised Learning

regression

continuous variable prediction



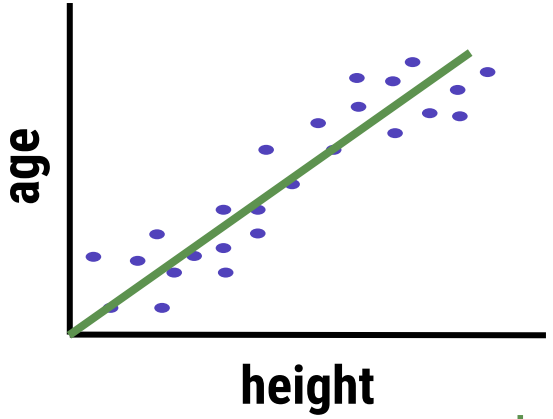


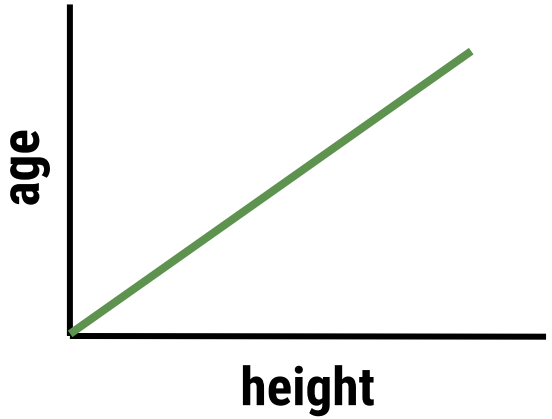
For prediction, the individual values in the training data are *not* important. We only need the model.

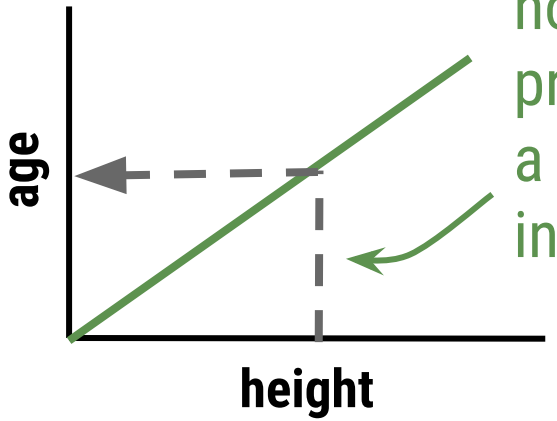
Supervised Learning

regression

continuous variable prediction





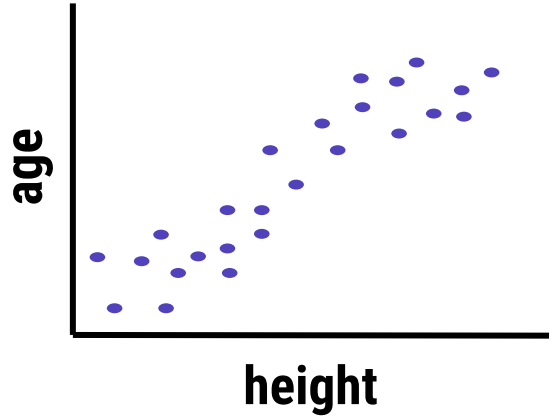


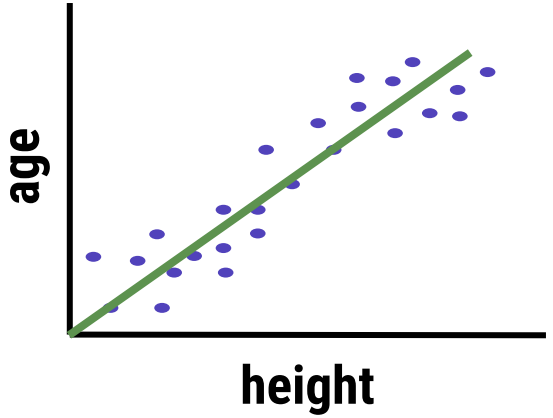
how we'll make predictions for a future individual

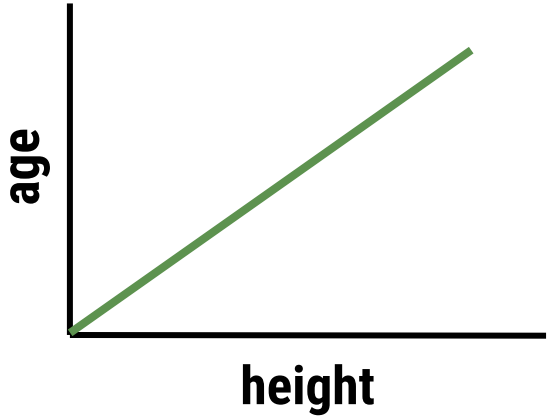
Supervised Learning

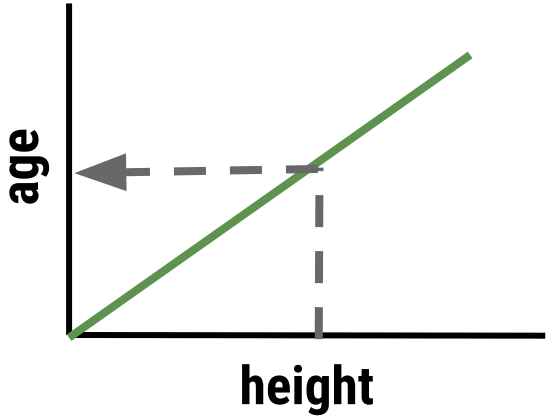
regression

continuous variable prediction

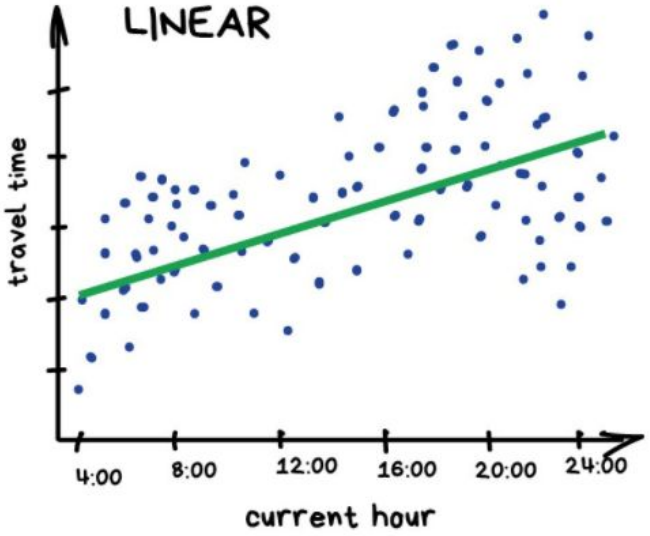


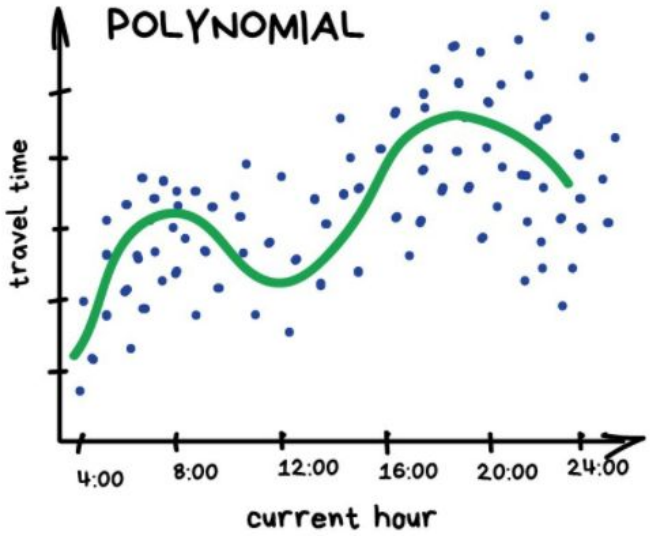






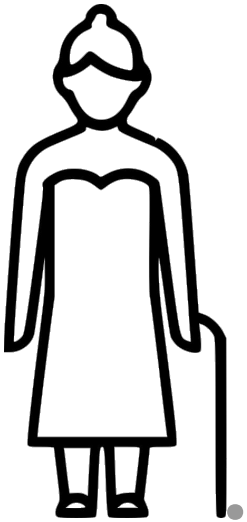
PREDICT TRAFFIC JAMS





REGRESSION

Image source: https://vas3k.com/blog/machine_learning/



Regression:

predicting continuous variables
(i.e. Age)

Classification:

predicting categorical variables
(i.e. education level)

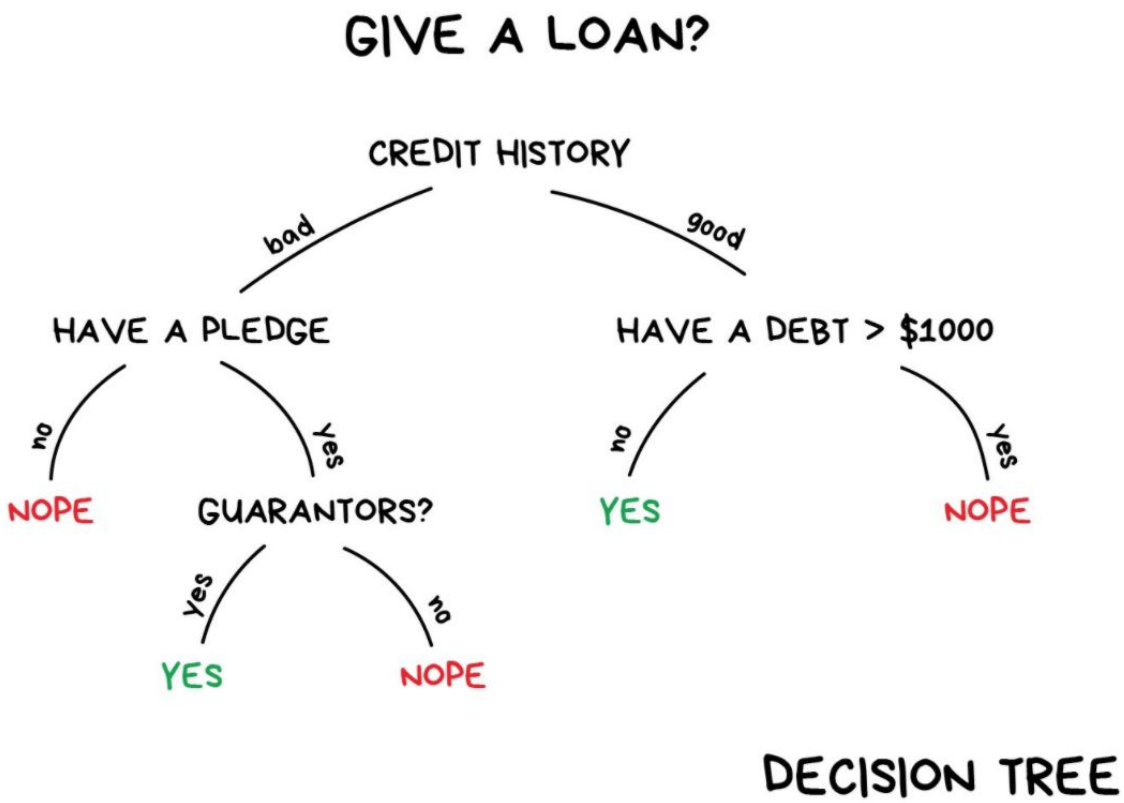


Image source: https://vas3k.com/blog/machine_learning/

Supervised Learning

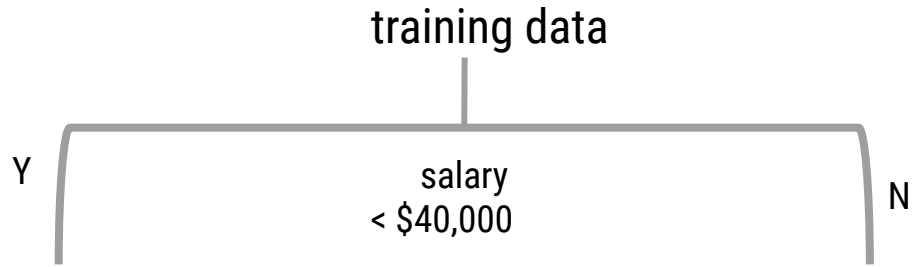
classification

categorical variable prediction

Supervised Learning

classification

categorical variable prediction



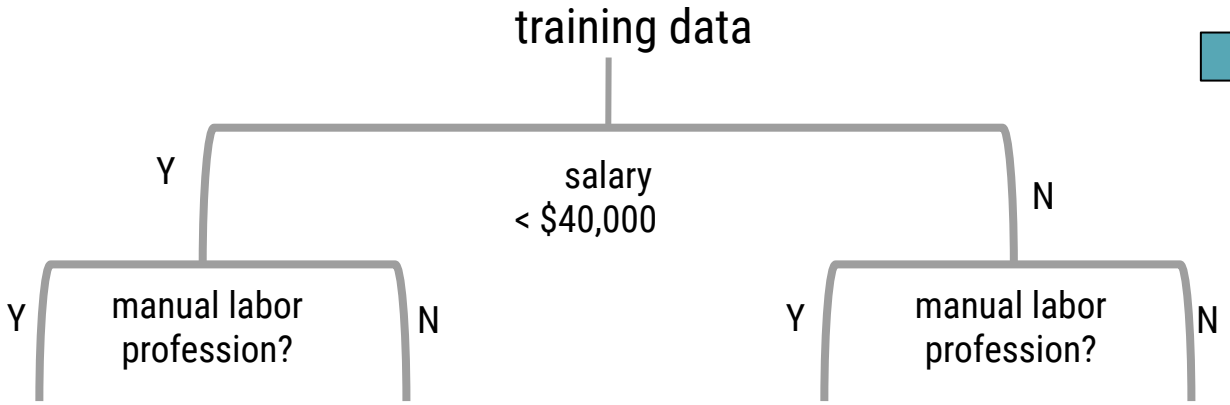
All the people who make *less* than 40K over here

All the people who make *more* than 40K over here

Supervised Learning

classification

categorical variable prediction

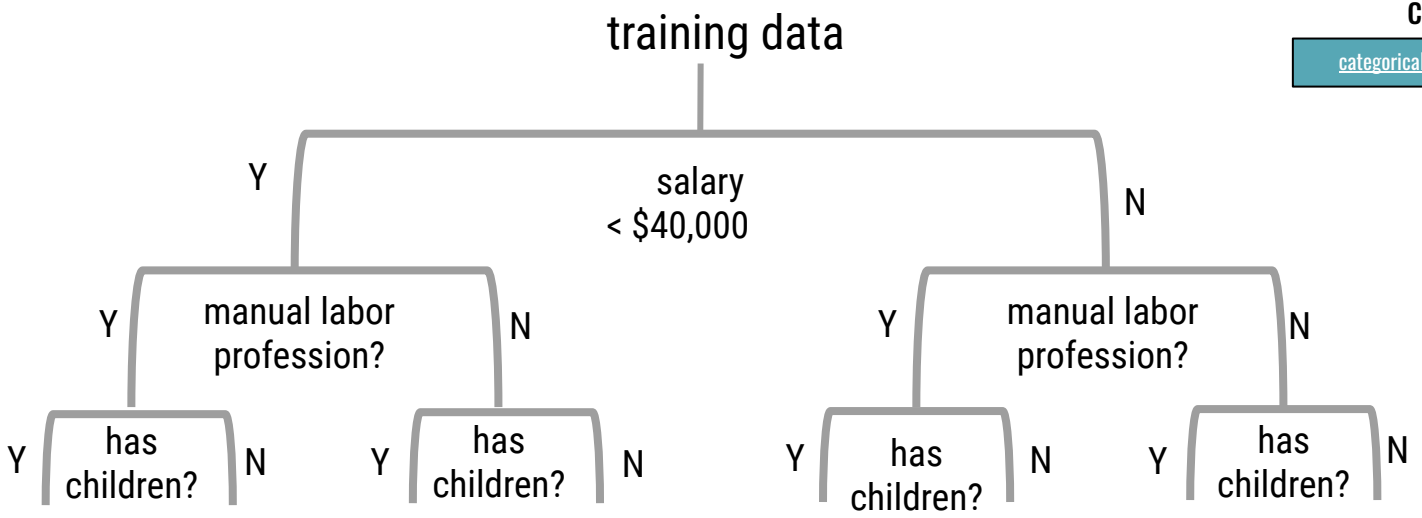


Continue adding *branches* to the **decision tree** where the variables and information in the training data decide which observations goes down which branch

Supervised Learning

classification

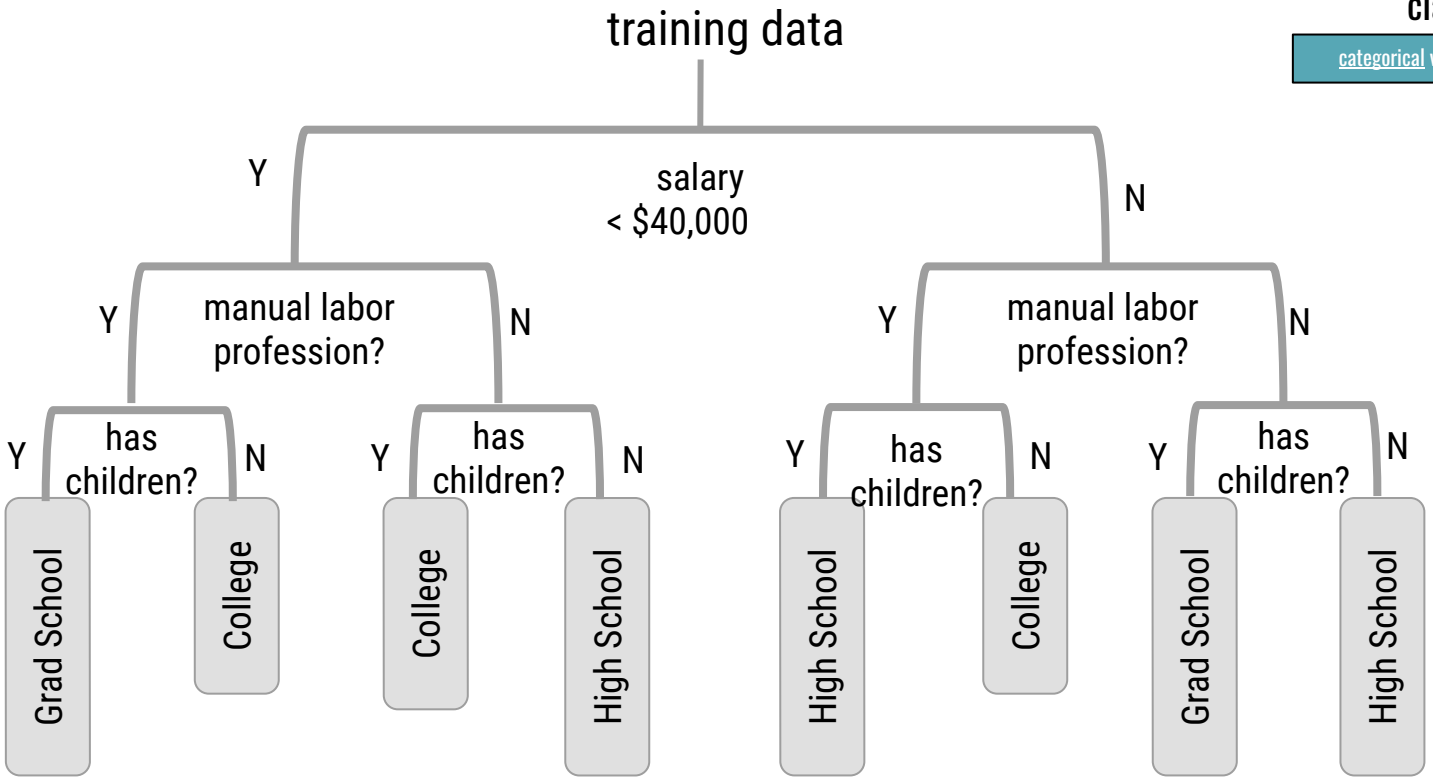
categorical variable prediction



Supervised Learning

classification

categorical variable prediction

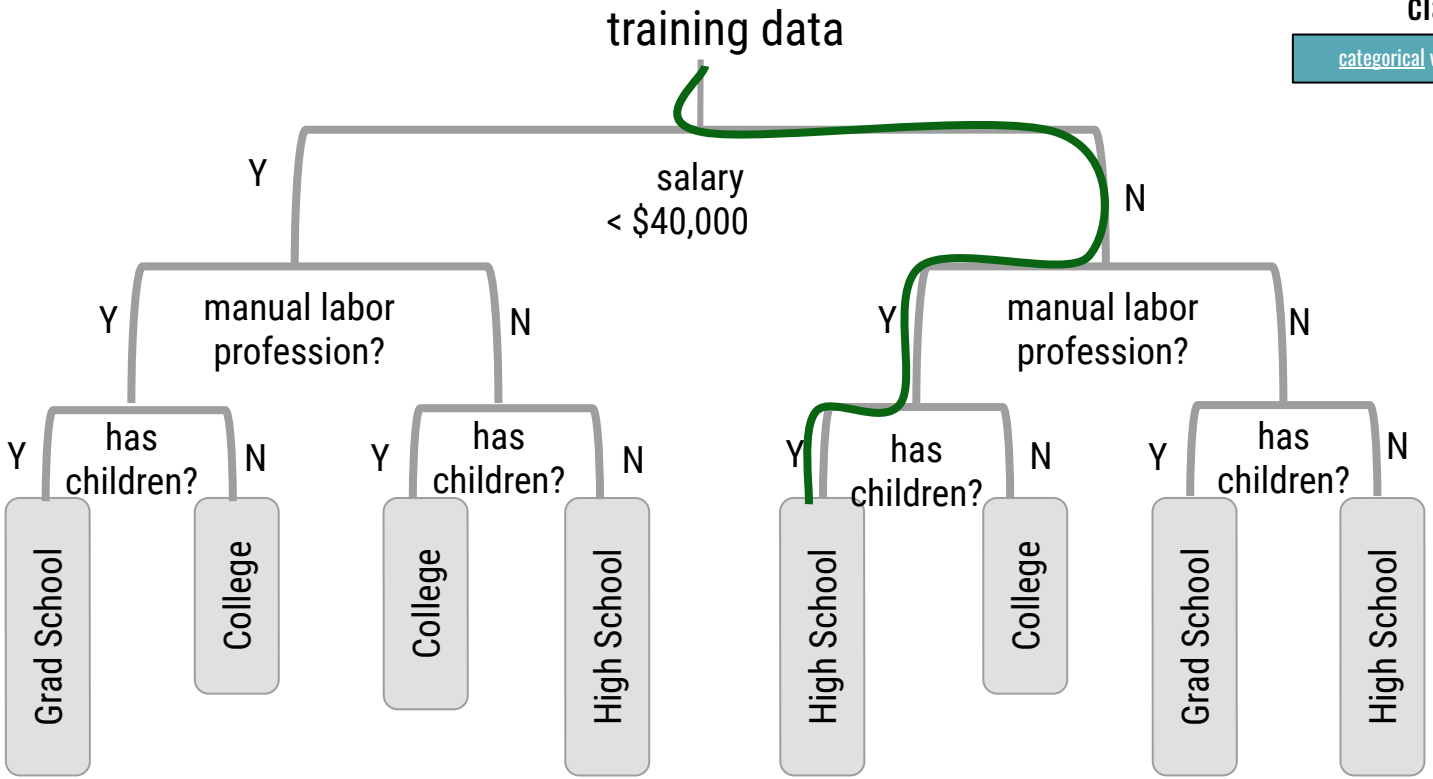


At the end of the tree, labels will be applied to each *leaf* of the tree

Supervised Learning

classification

categorical variable prediction

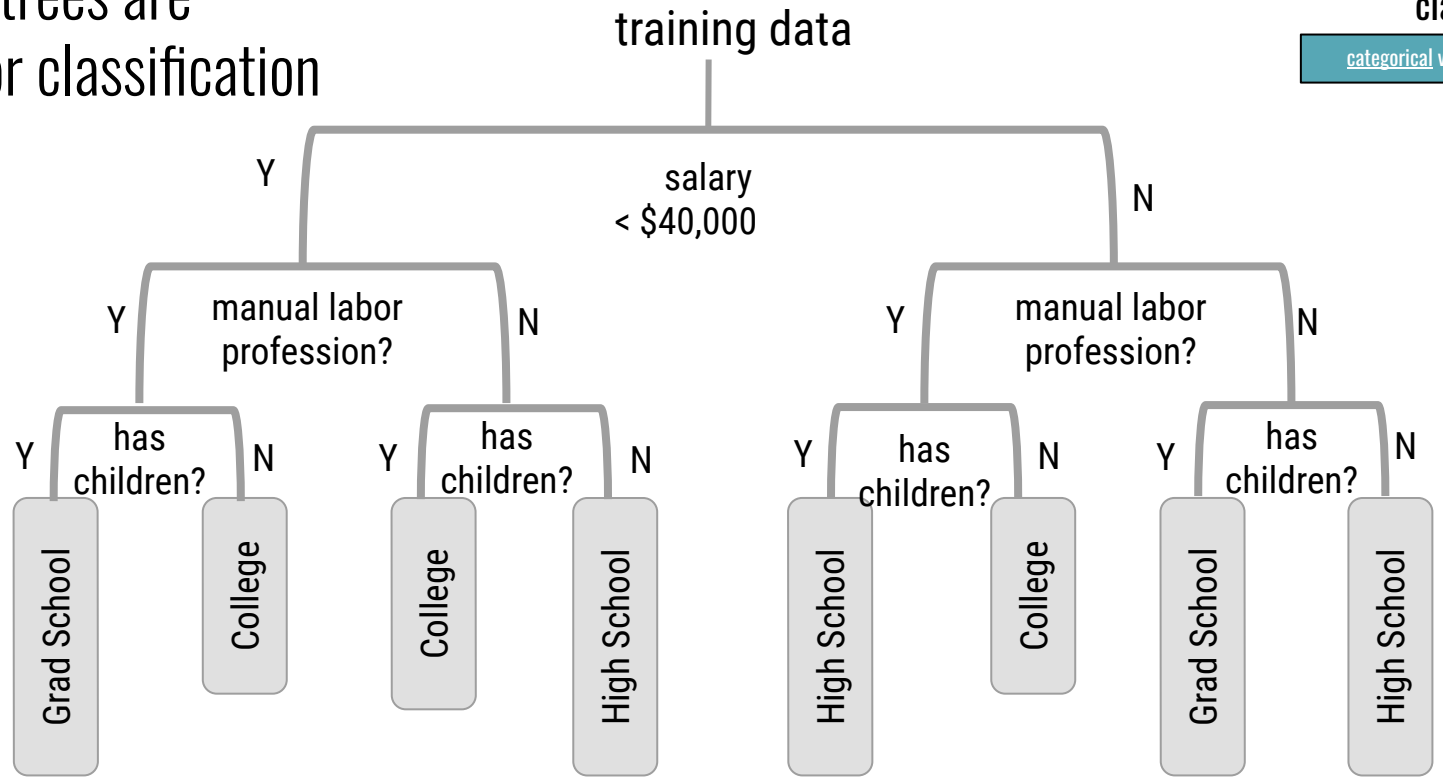


Decision trees are helpful for classification

Supervised Learning

classification

categorical variable prediction



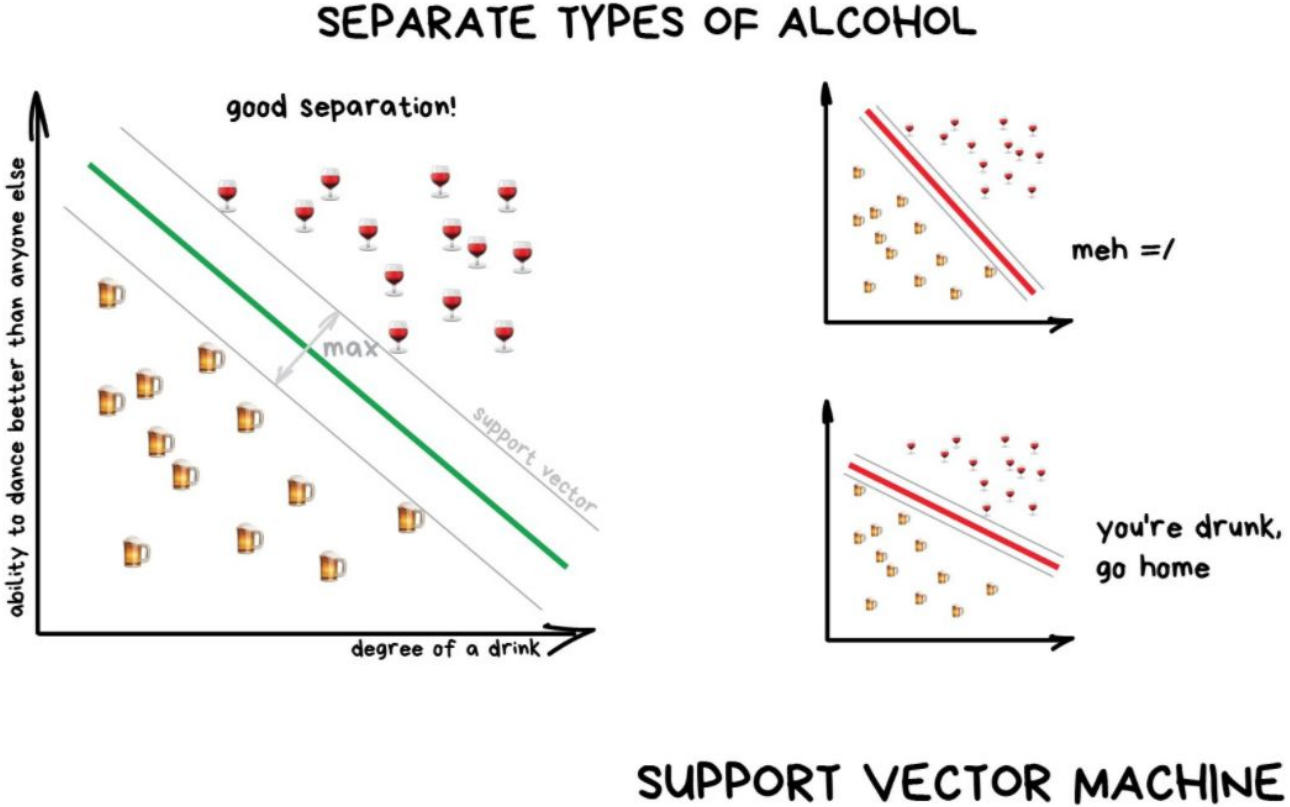
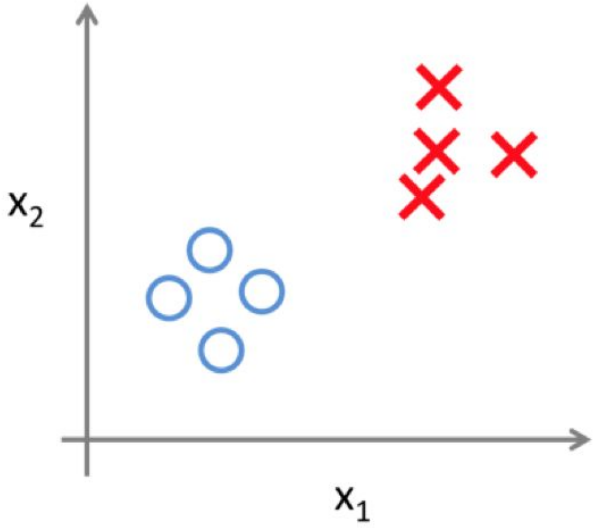


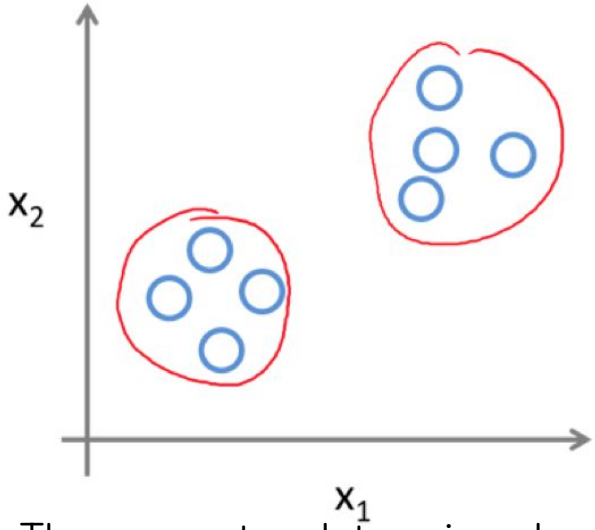
Image source: https://vas3k.com/blog/machine_learning/

To modes of machine learning

Supervised Learning



Unsupervised Learning



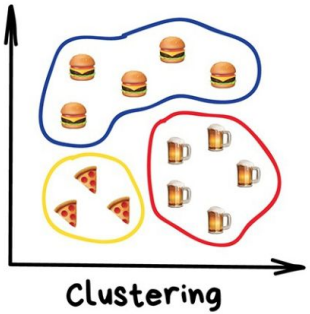
The computer determines how to classify based on properties within the data

Clustering

"Divides objects based on unknown features. Machine chooses the best way"

Nowadays used:

- For market segmentation (types of customers, loyalty)
- To merge close points on a map
- For image compression
- To analyze and label new data
- To detect abnormal behavior



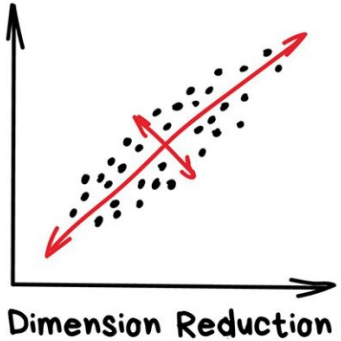
Popular algorithms: K-means_clustering, Mean-Shift, DBSCAN

Dimensionality Reduction (Generalization)

"Assembles specific features into more high-level ones"

Nowadays is used for:

- Recommender systems (★)
- Beautiful visualizations
- Topic modeling and similar document search
- Fake image analysis
- Risk management



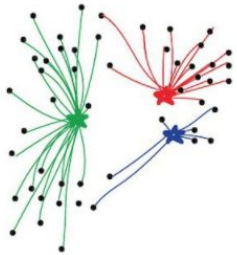
Popular algorithms: Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Latent Dirichlet allocation (LDA), Latent Semantic Analysis (LSA, pLSA, GLSA), t-SNE (for visualization)

Image source: https://vas3k.com/blog/machine_learning/

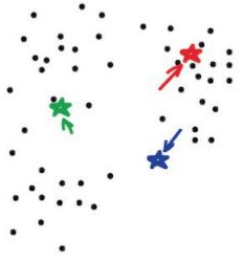
PUT KEBAB KIOSKS IN THE OPTIMAL WAY

(also illustrating the K-means method)

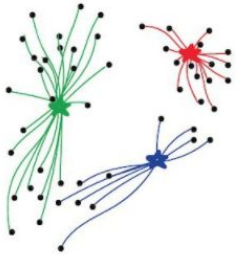
1. Put kebab kiosks in random places in city



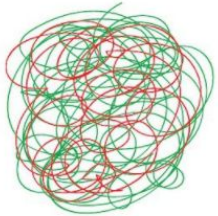
2. Watch how buyers choose the nearest one



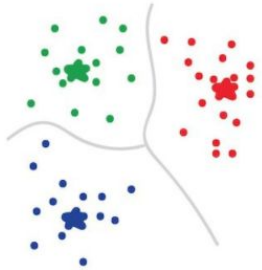
3. Move kiosks closer to the centers of their popularity



4. Watch and move again



5. Repeat a million times



6. Done!
You're god of kebabs!

Image source: https://vas3k.com/blog/machine_learning/