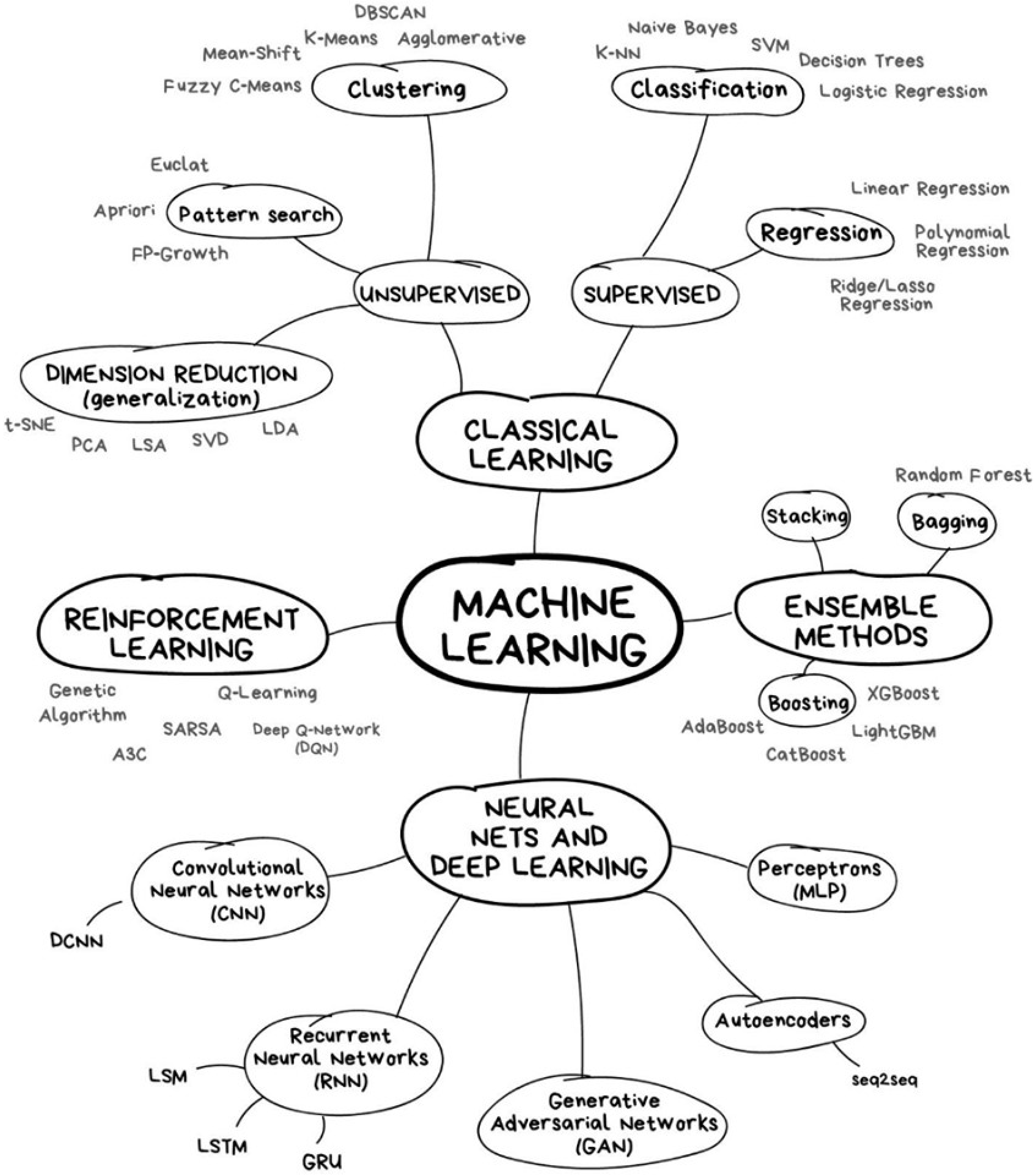
AI & Machine Learning Applications

Terminology

Machine Learning, Data Science, Data Mining, Data Analysis, Statistical Learning, Knowledge Discovery in Databases, Pattern Discovery.

Machine learning is an application of artificial **intelligence** (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.


MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
SUPERVISED
Clustering
DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Classification
Naive Bayes
K-NN
SVM
Decision Trees
Logistic Regression
Pattern search
Euclat
Apriori
FP-Growth
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
REINFORCEMENT LEARNING
Genetic Algorithm
Q-Learning
SARSA
Deep Q-Network (DQN)
A3C
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seq2seq

Data everywhere!

- Google: processes 24 peta bytes of data per day.
- Facebook: 10 million photos uploaded every hour.
- Youtube: 1 hour of video uploaded every second.
- Twitter: 400 million tweets per day.
- Astronomy: Satellite data is in hundreds of PB.
- …
- \By 2020 the digital universe will reach 44 zettabytes...“

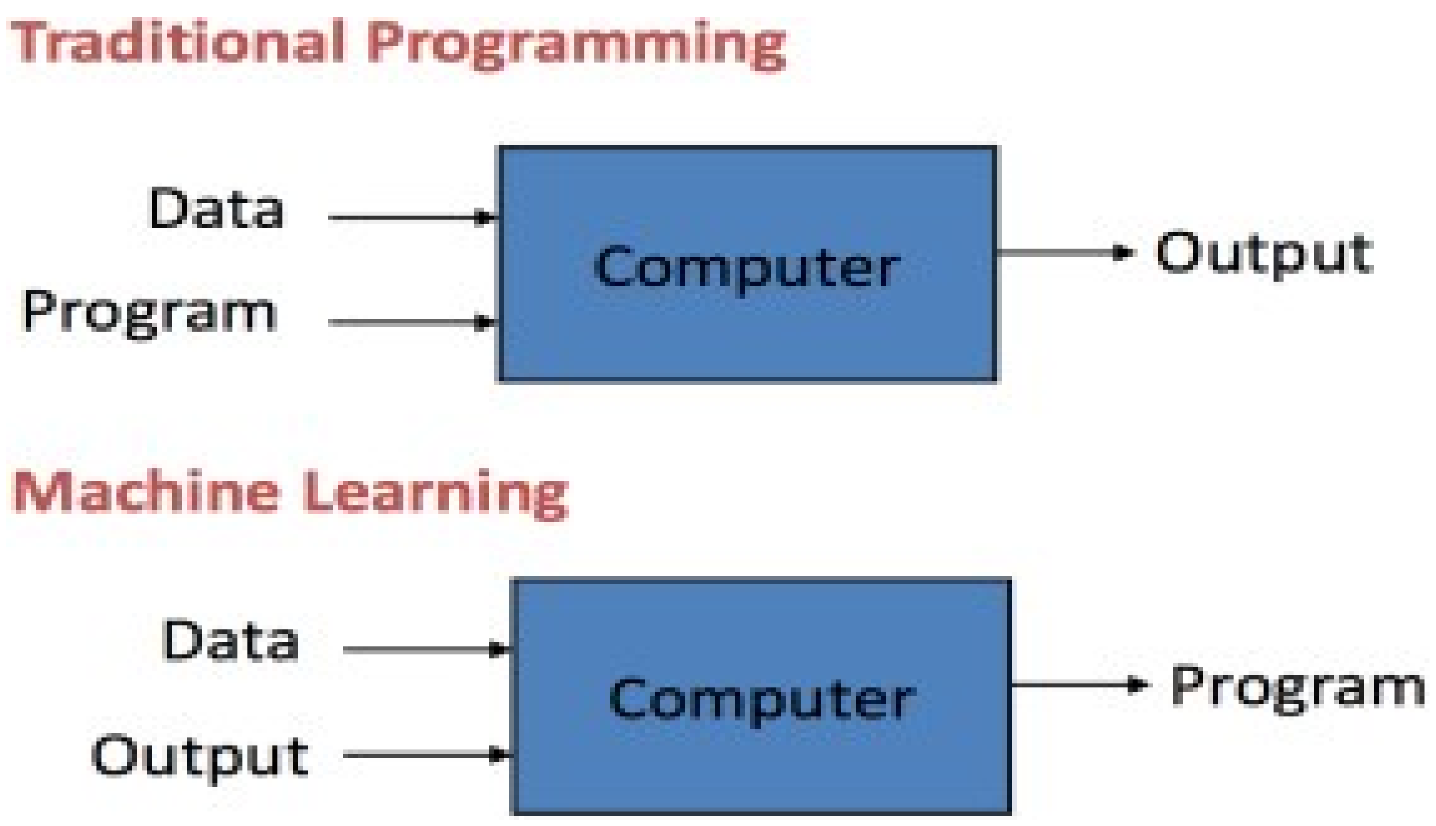
That's XYZ trillion gigabytes!

Data types

Data comes in different sizes and also flavors (types):

- Texts
- Numbers
- Clickstreams
- Graphs
- Tables
- Images
- Transactions
- Videos
- Some or all of the above!

Difference Between Traditional Programming and Machine Learning



Applications of Machine Learning

- Spam filtering
- Credit card fraud detection
- Digit recognition on checks, zip codes
- Detecting faces in images
- MRI image analysis
- Recommendation system
- Search engines
- Handwriting recognition
- Scene classification

Machine Learning:

- Decision trees
- Rule induction
- Neural Networks
- SVMs
- Clustering method
- Association rules
- Feature selection
- Visualization
- Graphical models
- Genetic algorithm

Key Elements of Machine Learning

Every machine learning algorithm has three components:

Representation: how to represent knowledge. Examples include decision trees, sets of rules, instances, graphical models, neural networks, support vector machines, model ensembles and others.

Evaluation: the way to evaluate candidate programs (hypotheses). Examples include accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence and others.

Optimization: the way candidate programs are generated known as the search process. For example combinatorial optimization, convex optimization, constrained optimization.

Feature Reduction in ML

- Theoretical view: More features More information

  More discrimination power

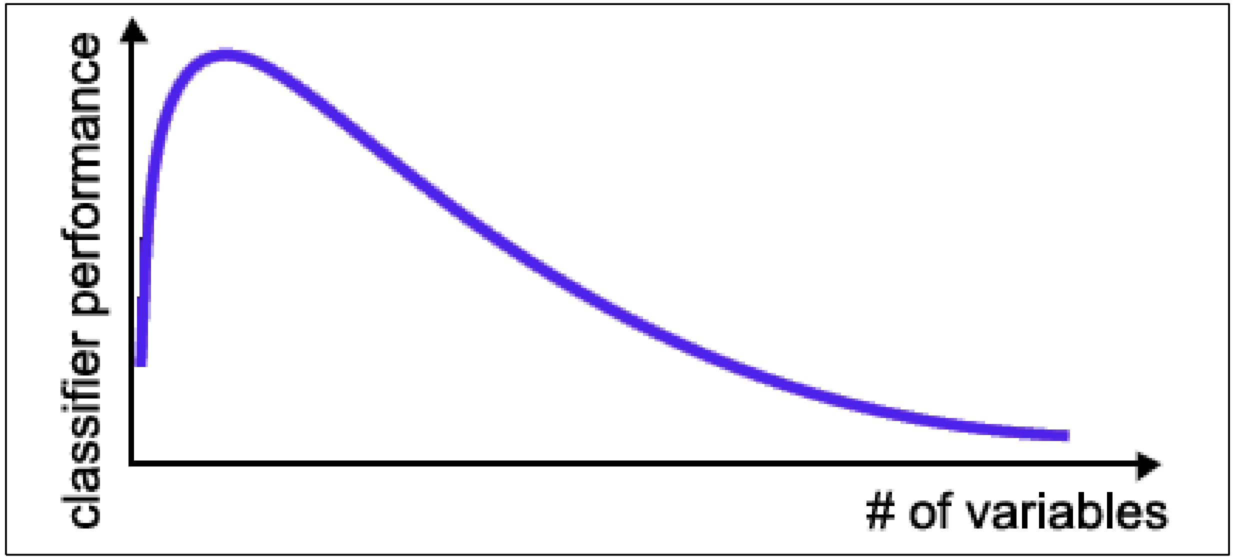
- In practice:
  - the inclusion of more features leads to worse performance

    many reasons why this is not the case!

Curse of Dimensionality

- number of training examples is fixed

  => the classifier's performance usually will degrade for a large number of features!



The number of training examples required increases exponentially with dimensionality.

Feature Reduction in ML

- Irrelevant and
- redundant features
  - can confuse learners.

- Limited training data.
- Limited computational resources.
- **Curse of dimensionality**.

Dimensionality Reduction

- Significant improvements can be achieved by first mapping the data into a *lower-dimensional* space.

$$x = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_N \end{bmatrix} --> reduce\ dimensionality --> y = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} \quad (K << N)$$

- Dimensionality can be reduced by:
  - Combining features using a linear or non-linear transformations.
  - Selecting a subset of features (i.e., feature selection).

Feature Selection

Problem of selecting some subset of features, while ignoring the rest

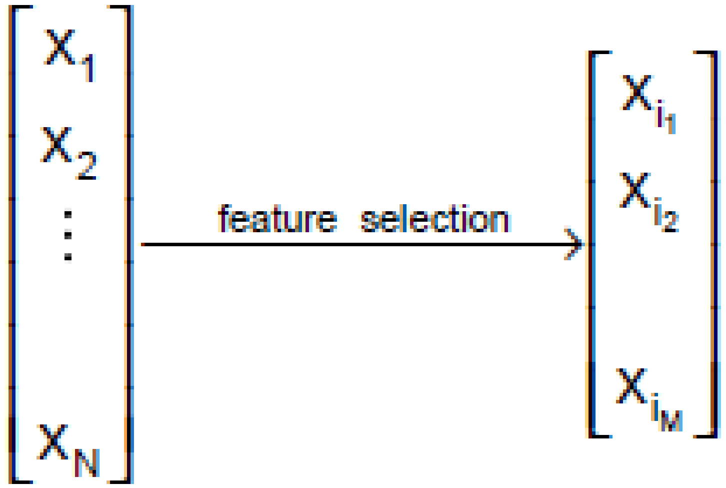
Feature Extraction

- Project the original x_i , i =1,...,d dimensions to new $k < d$ dimensions, z_j , j =1,...,k

Criteria for selection/extraction:
either improve or maintain the classification accuracy, simplify classifier complexity.

Feature Selection - Definition

- Given a set of features $F = \{x_1, \dots, x_n\}$
  the Feature Selection problem is
  to find a subset $F' \subseteq F$ that maximizes the learners ability to classify patterns.
- Formally F' should maximize some scoring function



Subset selection

- d initial features
- There are 2^d possible subsets
- Criteria to decide which subset is the best:
  - classifier based on these m features has the lowest probability of error of all such classifiers
- Can’t go over all 2^d possibilities
- Need some heuristics

Subset selection

- Select uncorrelated features
- Forward search
  - Start from empty set of features
  - Try each of remaining features
  - Estimate classification/regression error for adding specific feature
  - Select feature that gives maximum improvement in validation error
  - Stop when no significant improvement
- Backward search
  - Start with original set of size d
  - Drop features with smallest impact on error

Feature extraction - definition

- Given a set of features $F = \{x_1, \ldots, x_N\}$

  the Feature Extraction("Construction") problem is to map F to some feature set F'' that maximizes the learner's ability to classify patterns

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{bmatrix} \xrightarrow{\text{feature extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \\ y_M \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{bmatrix}\right)$$

Feature Extraction

- Find a projection matrix w from N-dimensional to M- dimensional vectors that keeps error low

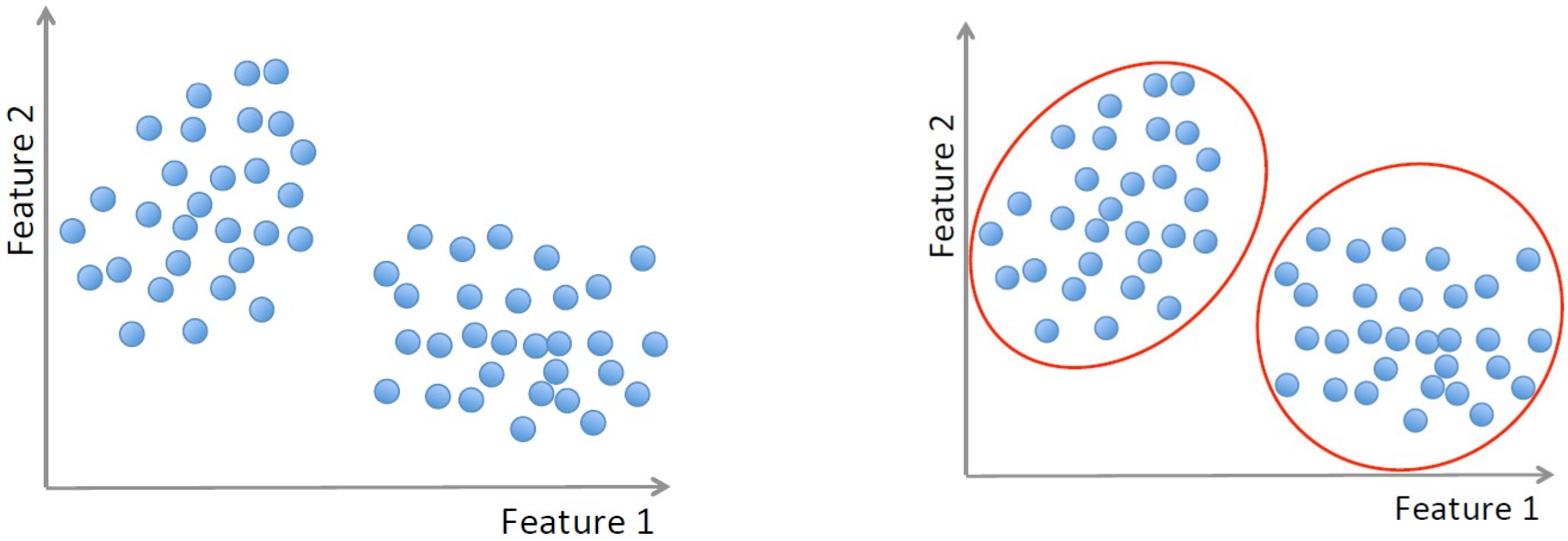
$$\boldsymbol{z} = w^T \boldsymbol{x}$$

Types of Learning

Unsupervised learning:

Learning a model from unlabeled data.

Unsupervised learning



Methods: K-means, gaussian mixtures, hierarchical clustering, spectral clustering, etc.

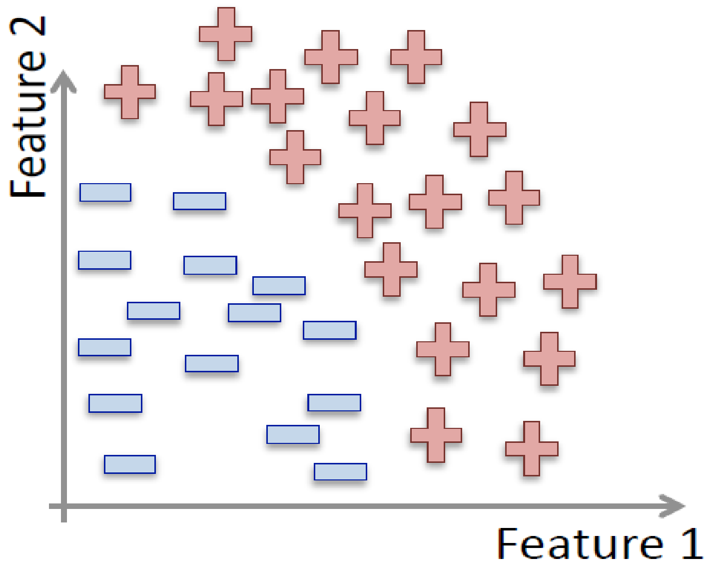
Types of Learning

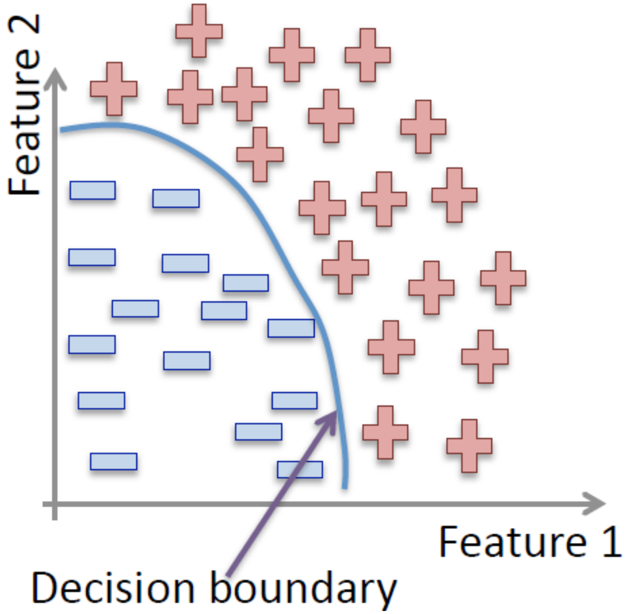
Supervised learning:

Learning a model from labeled data.

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
...				
fruit n	...	...	...	...

Supervised learning





Methods: Support Vector Machines, neural networks, decision trees, K-nearest neighbors, naive Bayes, etc.

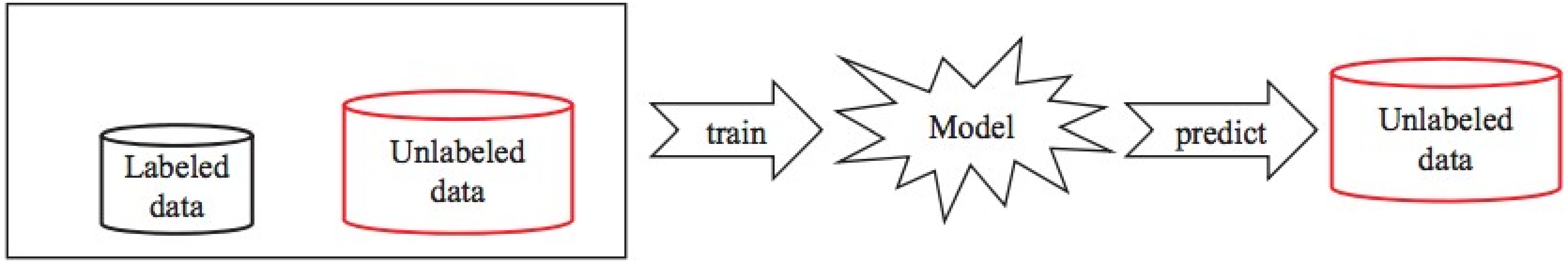
Types of Learning

Semi supervised learning:

Learning a model from unlabeled and labeled data.

(a) ImageNet Synset: One sample image from each category (b) Corel-1000 Dataset: Sample images from each category

(c) Caltech-256 Dataset: One sample image from each category (d) Caltech-101 Dataset: One sample image from each category



Types of Learning

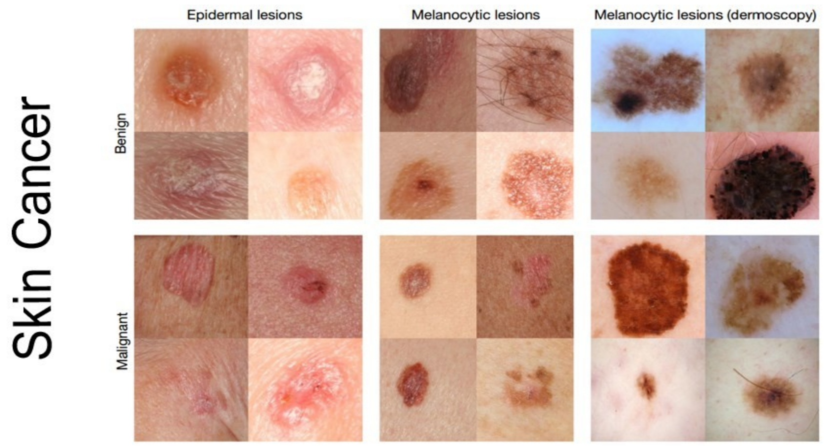
Weakly Supervised learning:

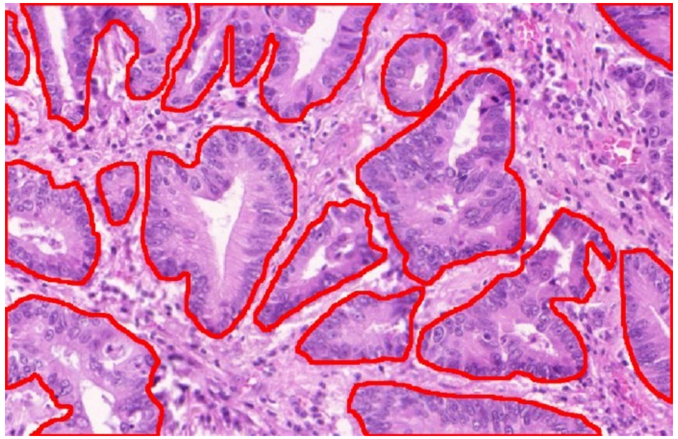
Learning a model from unlabeled and labeled labeled data.

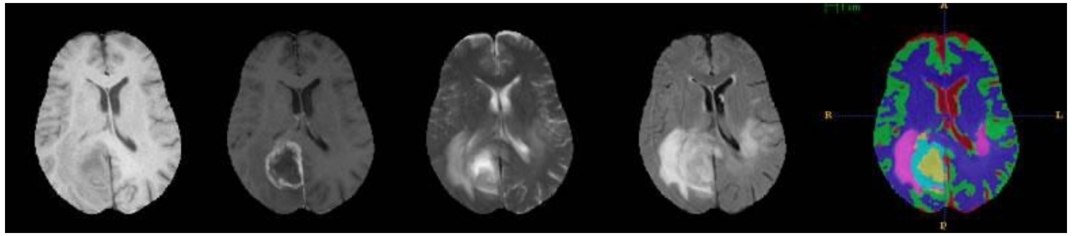
Caltech-UCSD Birds-200-2011 (CUB200-2011)

Some Applications

- Skin cancer detection
- Diabetics retinopathy detection
- Brain tumor segmentation
- Cell segmentation in microscopy images
- Cell segmentation and classification with partial annotations







Unsupervised learning

Introduction

- Supervised classification Algorithms is called as Classification.

- Unsupervised classification Algorithms is called as Clustering.

- Pattern is called as feature vector or observation.

- Depending on application we select the features.

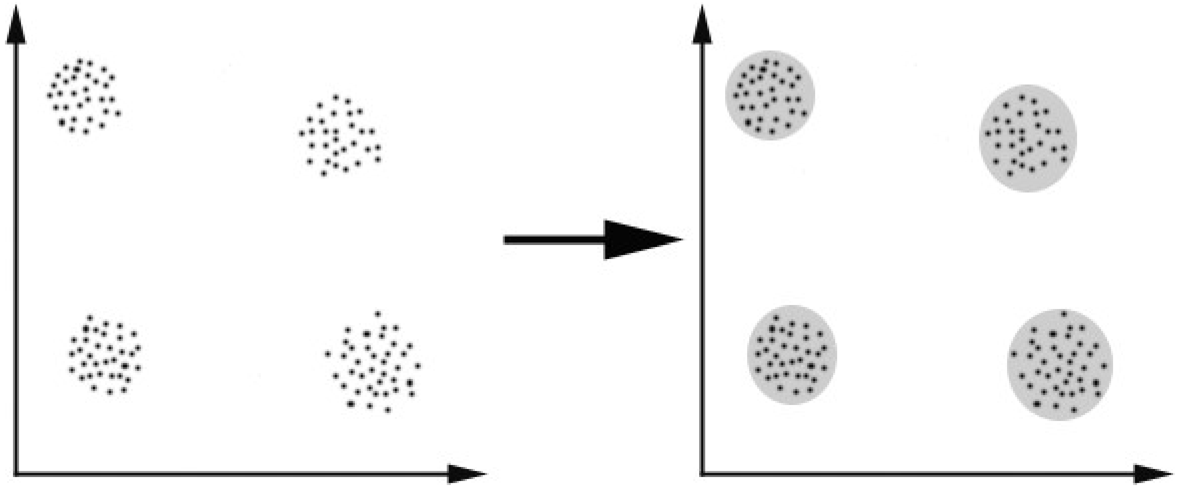
Examples:

Text Documents: Stop words, keywords, etc.
Video :color, color histogram, image edges, etc.
Finger Prints : loops, arch, ridges, valleys, etc.

Introduction

Clustering :

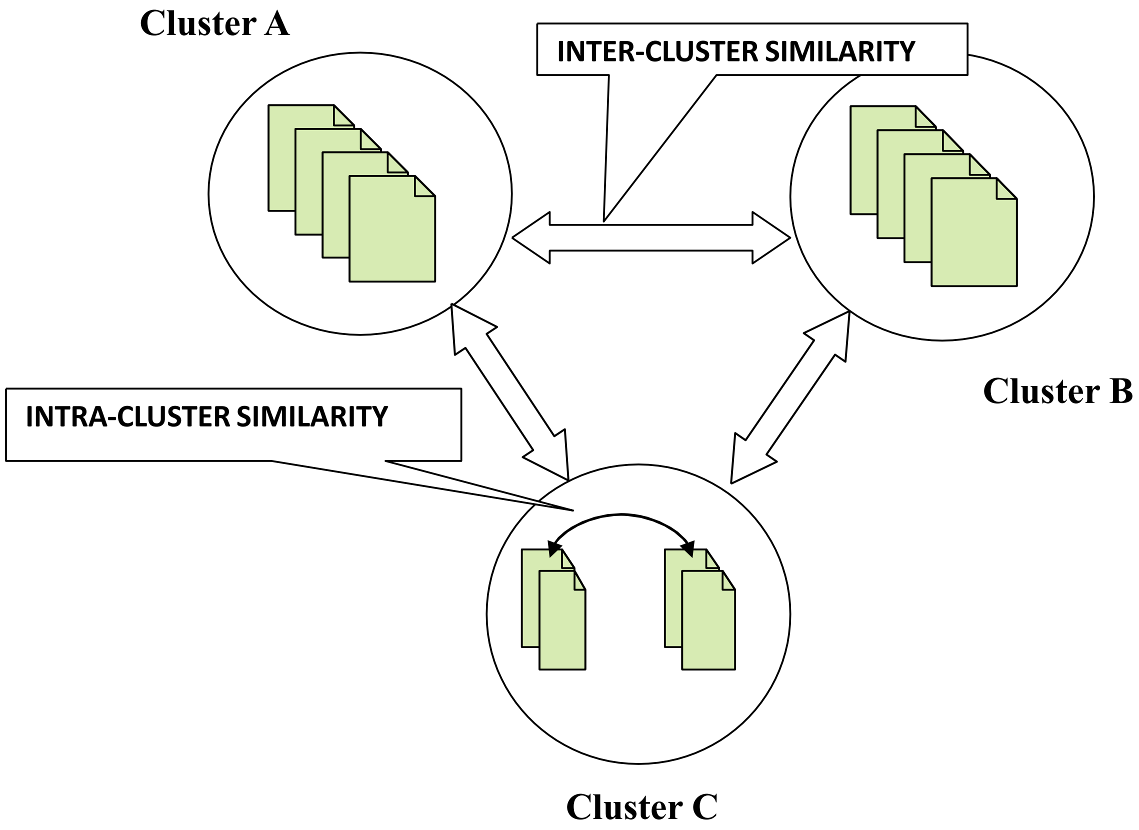
Clustering is the process of grouping a set of data objects into groups of similar objects.



Clustering Background

Cluster : A collection of data objects

- similar to one another within the same cluster
- dissimilar to objects in other clusters



Clustering Background

- One of the functionalities of pattern recognition.
- Unsupervised learning
- Different to classification
- No training set required
- Applications : Marketing,
  Biology,
  Documents.

Uses: Clustering useful in several exploratory Pattern analysis, Grouping Decision makings and Machine learning algorithms.

Challenges in Clustering

- Clustering of **Large data set.**
- **Fastness** of Clustering Algorithm.

Examples of Clustering Applications

- Marketing: Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs

- Land use: Identification of areas of similar land use in an earth observation database
- Insurance: Identifying groups of motor insurance policy holders with a high average claim cost

- City-planning: Identifying groups of houses according to their house type, value, and geographical location

- Earth-quake studies: Observed earth quake epicenters should be clustered along continent faults

Quality: What is Good Clustering?

- A good clustering method will produce high quality clusters with
  - high intra-class similarity
  - low inter-class similarity
- The quality of a clustering result depends on both the similarity measure used by the method and its implementation

Major Clustering Approaches

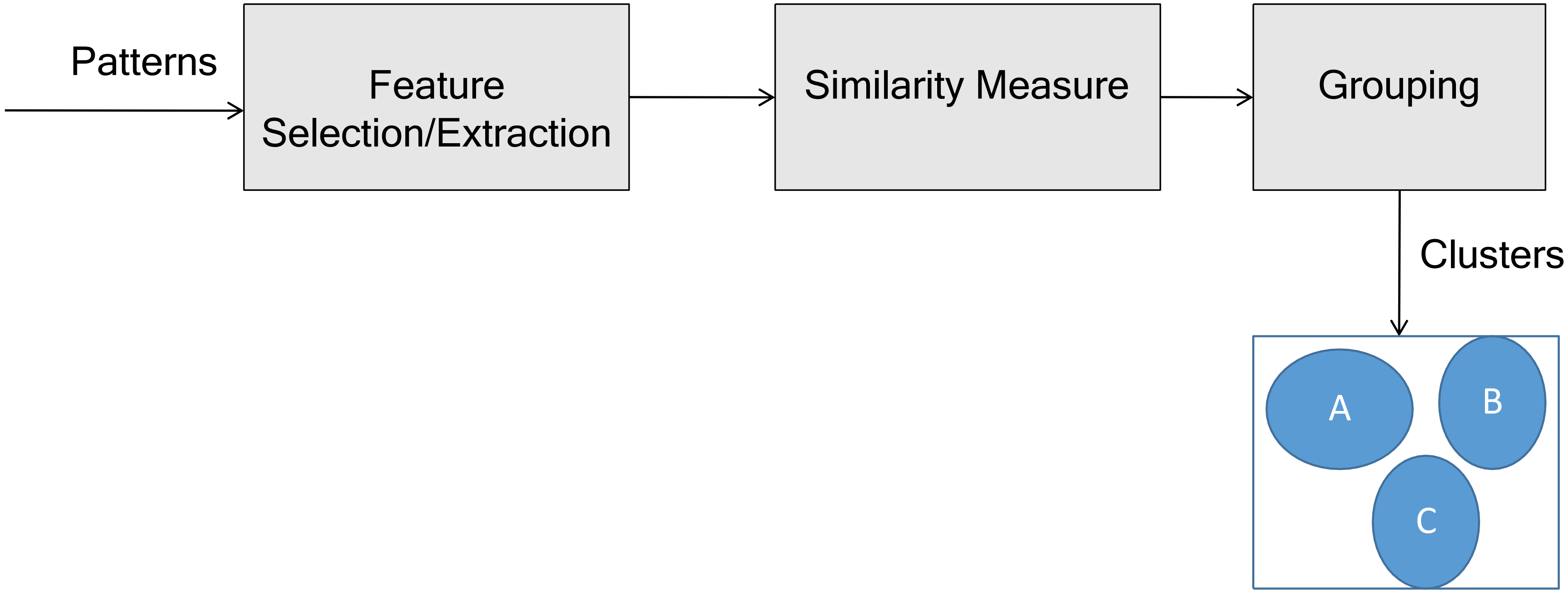
- Partitioning approach:
  - Construct various partitions and then evaluate them by some criterion, e.g., minimizing

    the sum of square errors
  - Typical methods: k-means, k-medoids, CLARANS
- Hierarchical approach:
  - Create a hierarchical decomposition of the set of data (or objects) using some criterion
  - Typical methods: Diana, Agnes, BIRCH, ROCK, CAMELEON
- Density-based approach:
  - Based on connectivity and density functions
  - Typical methods: DBSACN, OPTICS, DenClue

Major Clustering Approaches

- Grid-based approach:
  - based on a multiple-level granularity structure
  - Typical methods: STING, WaveCluster, CLIQUE
- Model-based:
  - A model is hypothesized for each of the clusters and tries to find the best fit of that model to each other
  - Typical methods: EM, SOM

System Architecture



Typical Alternatives to Calculate the Distance between Clusters

- Single link: smallest distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = min(t_{ip}, t_{jq})$
- Complete link: largest distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = max(t_{ip}, t_{jq})$
- Average: avg distance between an element in one cluster and an element in the other, i.e., $dis(K_i, K_j) = avg(t_{ip}, t_{jq})$
- Centroid: distance between the centroids of two clusters, i.e., $dis(K_i, K_j) = dis(C_i, C_j)$
- Medoid: distance between the medoids of two clusters, i.e., $dis(K_i, K_j) = dis(M_i, M_j)$
- Medoid: one chosen, centrally located object in the cluster

Partitioning Clustering Algorithms

Partitioning Method: Given a database of n objects, a portioning method constructs k partitions of the data, where each partition represents a cluster and k<n

Requirements:

1. Each group must contains at least one object.
2. Each object must belongs to exactly one group.

- *k-means* (MacQueen'67): Each cluster is represented by the center of the cluster
- *k-medoids* or PAM (Partition around medoids) (Kaufman & Rousseeuw'87): Each cluster is represented by one of the objects in the cluster

K-Means (Centroid -Based Technique)

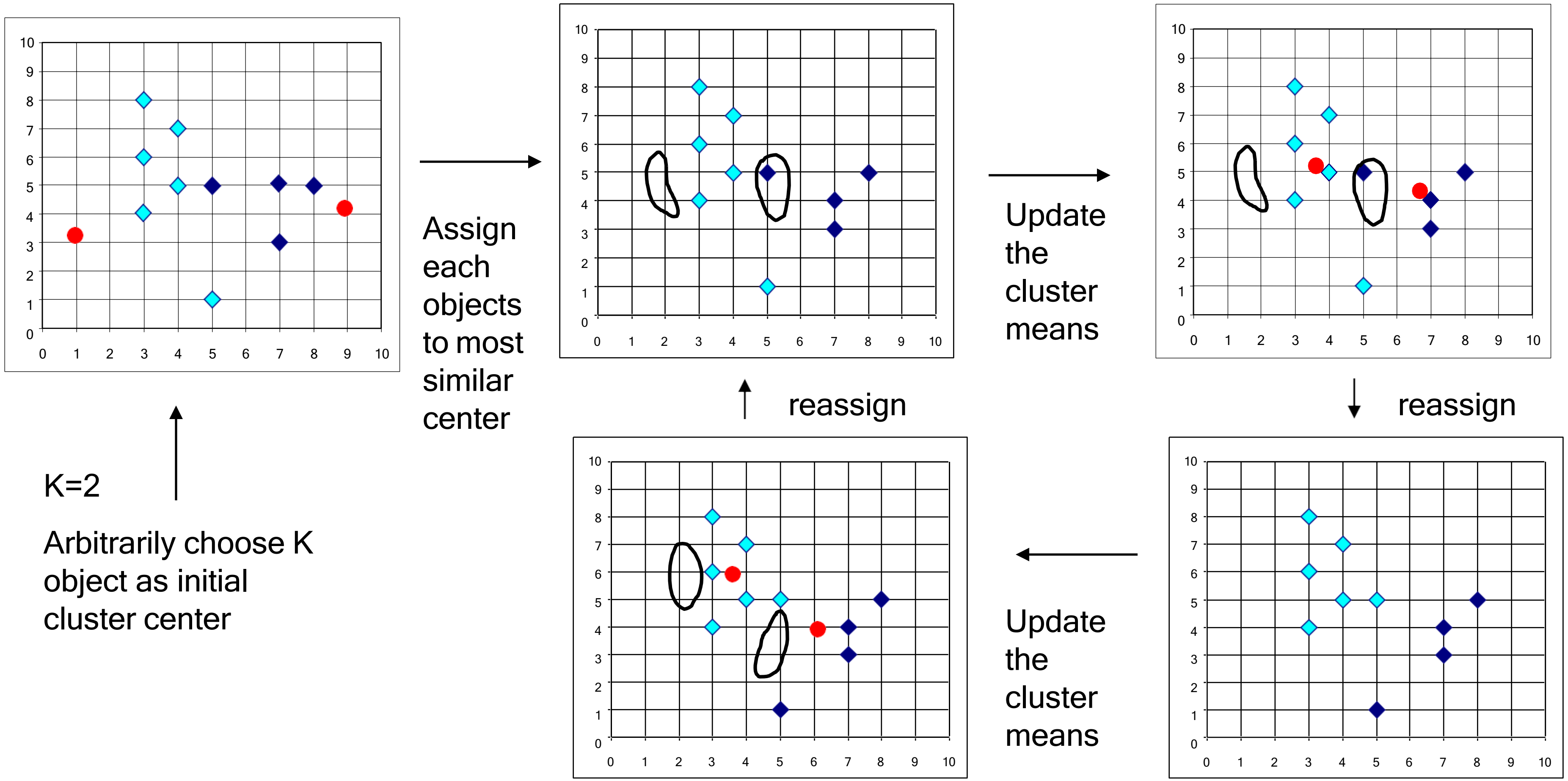
Input: The number of clusters k, data base contain n objects.

Output: A set of k clusters. Method:

1. arbitrarily choose k objects as the initial cluster centers.
2. repeat
3. assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.
4. update the cluster center means.
5. until no change.

The *K-Means* Clustering Method

- Example



Drawbacks

- ❖Applicable only when *mean* is defined
- ❖Need to specify *k,* the *number* of clusters, in advance
- ❖Unable to handle noisy data and *outliers*
- ❖Not suitable to discover clusters with *non-convex shapes*

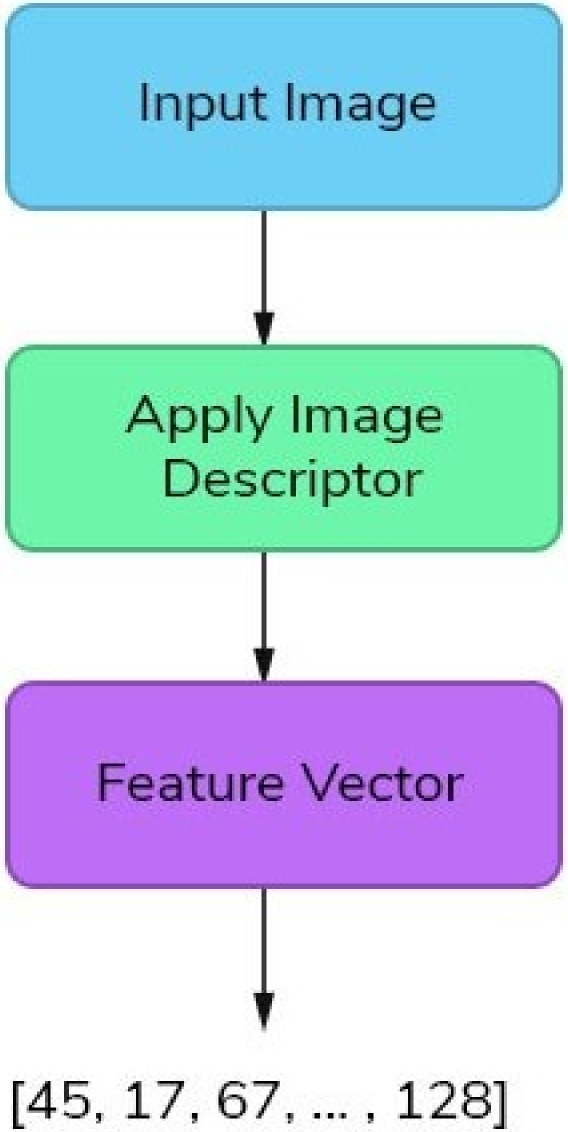
Applications Related to Mechanical Dept.

- Classification methods are used to find out the likelihood of new shaft being failure one under given load condition.

- Engine Prognosis

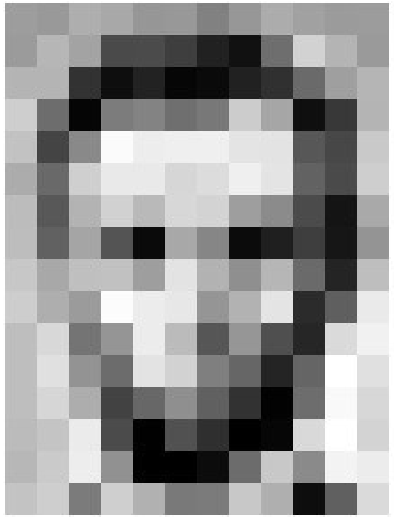
- Tool condition monitoring (using vibration/sound/acceleration data to predict wear and tear through regression)

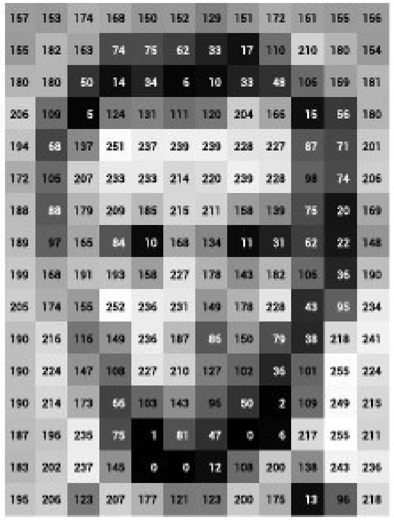
Image Classification


Input Image
Apply Image Descriptor
Feature Vector
[45, 17, 67, ... , 128]

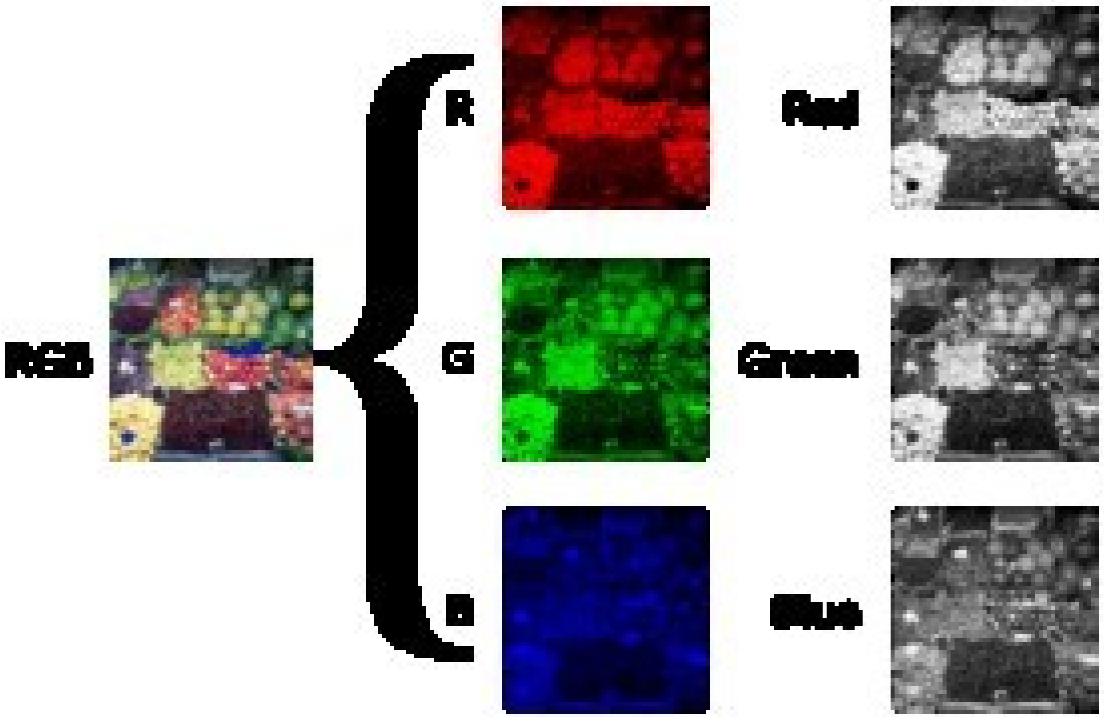
Feature Extraction

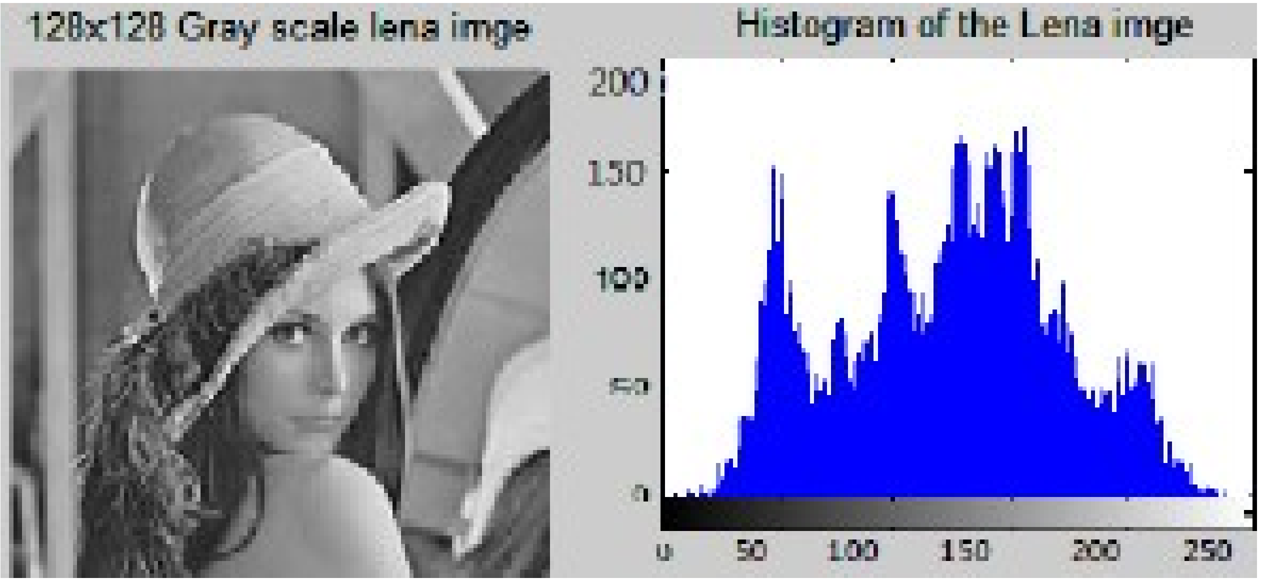
Features are the information or list of numbers that are extracted from an image.
These are real-valued numbers (integers, float or binary).



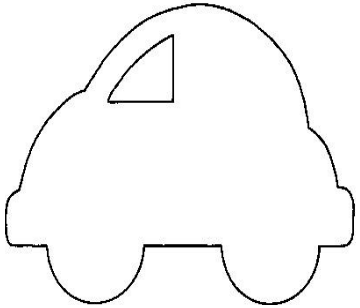


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218




128x128 Gray scale lena imge
Histogram of the Lena imge
200
150
100
50
0
0
50
100
150
200
250

Shape



Texture

Global Feature Descriptors

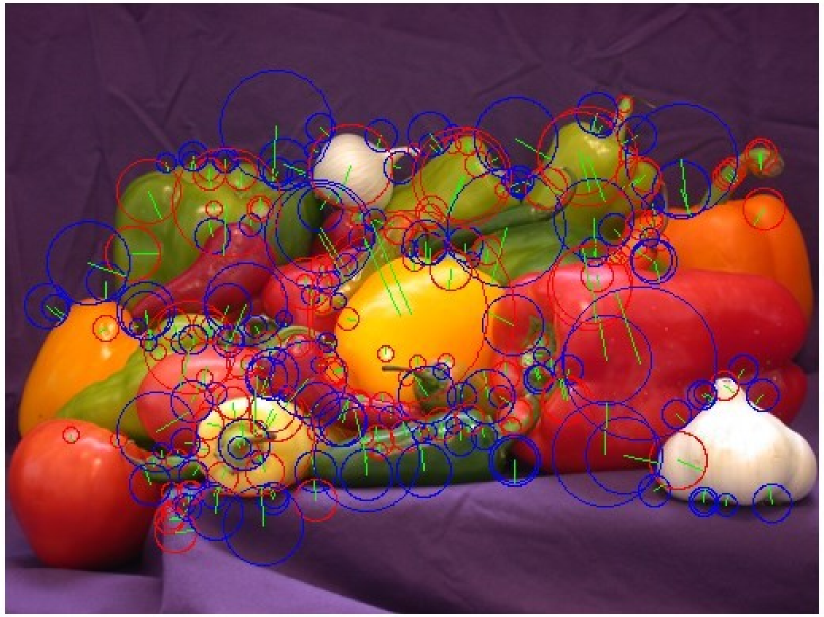
These are the feature descriptors that quantifies an image globally.

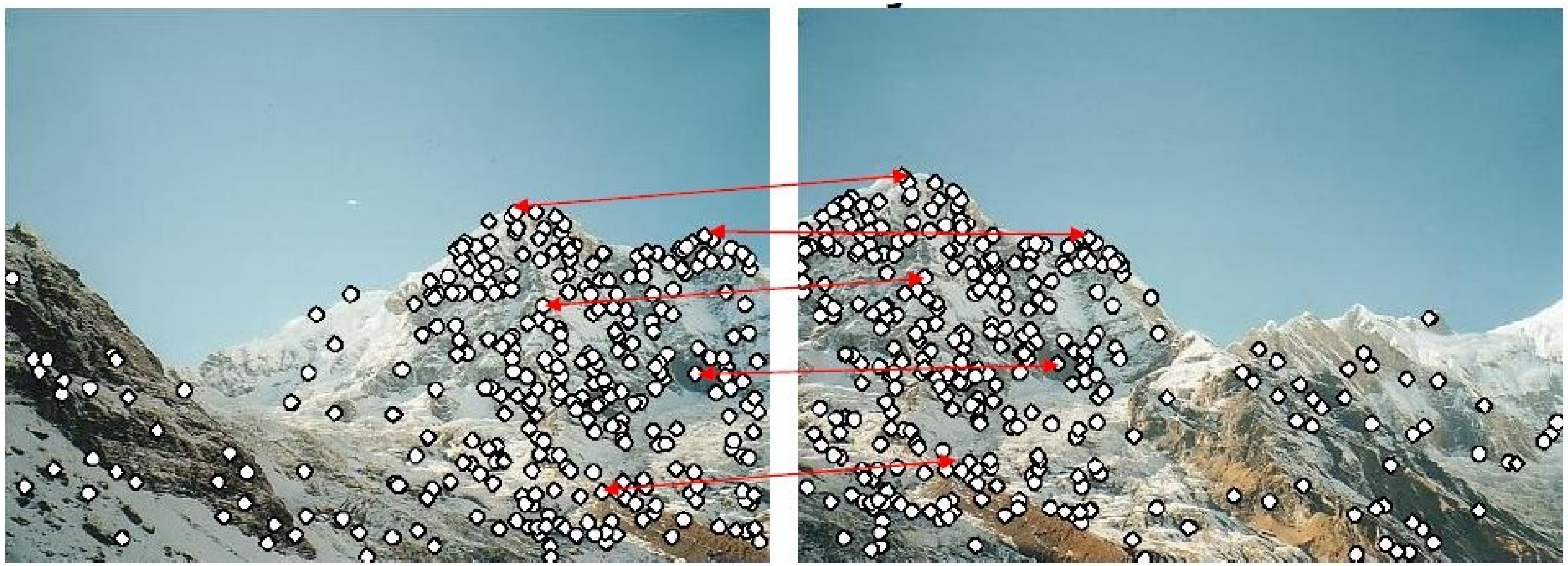
- **Color** - Color Channel Statistics (Mean, Standard Deviation) and Color Histogram
- **Shape**
- **Texture**
- **Others** - Histogram of Oriented Gradients (HOG)

Local Feature Descriptors

- These are the feature descriptors that quantifies local regions of an image.
- Interest points are determined in the entire image and image patches/regions surrounding those interest points are considered for analysis.

- SIFT (Scale Invariant Feature Transform)
- SURF (Speeded Up Robust Features)
- BRIEF (Binary Robust Independed Elementary Features)





FLOWERS-17 dataset

Thank You