Introduction

[B] batch learning

[C] curse of dimensionality

[D] decision boundary

[F] feature / attribute

[G] generalization

[L] label

[M] machine learning

model / hypothesis

[O] Occam's razor

online learning

overfitting

[P] prediction

predictor

[R] reinforcement learning

[S] semi-supervised learning

supervised learning

[U] underfitting

unsupervised learning

Decision Theory

[B] Bayes optimal

Bayes' theorem

[C] class conditional probability

class prior probability

conditionally independent

error minimization [E]

Laplace smoothing [L]

loss function

naïve Bayes [N]

risk minimization [R]

scaling [S]

Density Estimation

Bayesian information criterion (BIC) [B]

binning

EM algorithm [E]

Gaussian distribution [G]

Gaussian kernel

Gaussian mixture model

histogram [H]

independent and identically distributed (iid) [I]

kernel density estimation [K]

latent variable [L]

likelihood

log-likelihood

maximum likelihood estimation [M]

mixture model

multivariate Gaussian distribution

nearest neighbors [N]

non-parametric models

observed variable [O]

P parametric models

Parzen window

Regression

B basis function (polynomial, RBF)

batch gradient descent

D data input space

design matrix / regressor matrix

E elastic net

F feature mapping

G gradient descent

H hypothesis space

L Lasso regression

learning rate (α)

M mini-batch gradient descent

N noise

normal equation

O ordinary least squares (OLS)

P precision

probabilistic regression

R regression

regularization

ridge regression

S squared error

stochastic gradient descent

Classification

activation (function) A

AdaBoost

average entropy

backpropagation B

bagging

batch-normalization

binary classification

boosting

classification and regression trees (CART) C

convolutional neural network (CNN)

cosine similarity

cross entropy

decision tree D

deep learning

distance metric

ensemble methods E

entropy

epoch

Euclidean distance

ExtraTrees

fully connected neural network F

Gain ratio (GR) G

Gini index

hidden layer H

hinge loss

Evaluation

train split

[V] variance

Unsupervised Learning

[A] affinity-based clustering

agglomerative clustering

[C] centroid

cluster

clustering

complete linkage

[D] dendrogram

dimensionality reduction

[E] eigenvalue problem

elbow method

EM-based clustering

[H] hierarchical clustering

[K] k-means

[M] maximum variance formulation

[O] orthogonal projection

[P] principal component analysis (PCA)

[S] single linkage

[V] vector quantization

voronoi diagram