

Introduction

		error minimization	E
^B	batch learning	Laplace smoothing	L
^C	curse of dimensionality	loss function	
^D	decision boundary	naïve Bayes	N
^F	feature / attribute	risk minimization	R
^G	generalization	scaling	S
^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		

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

^I	ID3	tangent hyperbolic (tanh)	^T
	information gain (IG)	word2vec	^W
	intrinsic information (IntI)		
^K	<i>k</i> -nearest neighbors	Evaluation	
^L	lazy learning		
^M	Manhattan distance	accuracy	^A
	Minkowski distance	area-under-the-curve (AUC)	
	multi-class classification	bias	^B
	Multi-layer perceptron (MLP)	confusion matrix	^C
^N	negative log-likelihood	cross-validation (X-Val)	
	neural network	dev split	^D
^O	one-hot encoding	early stopping	^E
	one-vs-one (OvO)	F_1 -score	^F
	one-vs-rest (OvR)	grid search	^G
^P	Perceptron	leave-one-out cross-validation (LOO X-Val)	^L
	Perceptron convergence theorem	macro average	^M
	pooling	mean absolute error (MAE)	
	pre-activation	micro average	
^R	random forest	out-of-sample testing	^O
	regression trees	precision	^P
	rectified linear unit (ReLU)	precision-recall curve	
	recurrent neural network (RNN)	random search	^R
^S	sigmoid function (σ)	recall	
	softmax	receiver operating characteristic (ROC)	
	stacking	root mean square error (RMSE)	
	standard deviation reduction (SDR)	stratified split	^S
		test split	^T

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