Definitions by Module

Last modified: 2026-11-21

Every definition in the notes, listed by module and page in the order the notes present them. Each row numbers the definition within its page and across the whole list, so the last overall number counts every definition. The list is built from the notes themselves, so it cannot fall out of step with them. The glossary explains the key terms in plain words and points to where each is defined. The results are listed the same way.

M1: Course Overview

Course Overview: What Is Machine Learning?

Table 1: The definitions on Course Overview: What Is Machine Learning?, in page order.

On page	Overall	Definition
1	1	Machine learning
2	2	Generalization
3	3	Generalizing well
4	4	Learning from experience
5	5	Artificial intelligence
6	6	Predicted value (prediction)
7	7	Prediction (predictive modeling)
8	8	Inference

Anatomy of a Learning Problem

Table 2: The definitions on Anatomy of a Learning Problem, in page order.

On page	Overall	Definition
1	9	Training data (training set)
2	10	Learning algorithm
3	11	Training (fitting, learning, estimation)
4	12	Feature (attribute)
5	13	Model
6	14	Statistical model
7	15	Predictive model
8	16	Feature vector
9	17	Feature function (feature map, encoding)
10	18	Output variable (label, target, response, outcome, dependent variable)
11	19	Structured data

On page	Overall	Definition
12	20	Unstructured data
13	21	Supervised learning
14	22	Unsupervised learning
15	23	Semi-supervised learning
16	24	Reinforcement learning
17	25	Clustering
18	26	k-means
19	27	Regression
20	28	Classification
21	29	Sample mean squared error
22	30	Sample root mean squared error
23	31	Evaluation function
24	32	Linear evaluation function
25	33	Backed-up score

M2: Linear Regression

Linear Regression

Table 3: The definitions on Linear Regression, in page order.

On page	Overall	Definition
1	34	Affine function
2	35	Linear function
3	36	Linear model
4	37	Level set
5	38	Contour plot
6	39	Dot product
7	40	Design vector (model vector, regressor vector)
8	41	The intercept trick (bias trick, 1-padding)
9	42	Regression task
10	43	Univariate regression
11	44	Residual
12	45	Squared error loss (square loss)
13	46	Ordinary least squares objective
14	47	The design matrix (feature matrix, model matrix, regressor matrix)
15	48	The gradient
16	49	Minimum-norm solution
17	50	The Moore–Penrose pseudoinverse
18	51	Condition number
19	52	Ill-conditioned matrix
20	53	Prediction at a new point
21	54	The hat matrix
22	55	Leverage
23	56	Sample mean squared error (sMSE)
24	57	Sample root mean squared error (sRMSE)

On page	Overall	Definition
25	58	Residual standard error (RSE, standard error of the regression)
26	59	Sum of squared errors (also called residual sum of squares)
27	60	Sample mean absolute error
28	61	Coefficient of determination (R^2, r^2)
29	62	Reducible error
30	63	Feature expansion, basis expansion
31	64	Linear in the parameters (linear functional form)
32	65	The Fourier basis expansion
33	66	Basis function

Interactions, Inference and Diagnostics

Table 4: The definitions on Interactions, Inference and Diagnostics, in page order.

On page	Overall	Definition
1	67	Interaction term (product term, cross term)
2	68	Hierarchy principle (hierarchical principle)
3	69	Standard error of a coefficient
4	70	Estimated standard error
5	71	Confidence interval for a coefficient
6	72	The t-test for a coefficient
7	73	The F-test that any feature is useful
8	74	The partial F-test for a group of features
9	75	Confidence interval for the mean response
10	76	Prediction interval
11	77	Residual plot
12	78	Standardized residual
13	79	Outlier
14	80	Variance inflation factor

Multivariate-Outcome Linear Regression

Table 5: The definitions on Multivariate-Outcome Linear Regression, in page order.

On page	Overall	Definition
1	81	Multivariate linear regression

M3: ML Fundamentals

A First Model, End to End

Table 6: The definitions on A First Model, End to End, in page order.

On page	Overall	Definition
1	82	Parameter (weight, coefficient)
2	83	Training set
3	84	Fitted model
4	85	Parametric model
5	86	True mean function
6	87	Noise conditional on the input
7	88	Signal-plus-noise decomposition
8	89	Loss function
9	90	Plausible loss
10	91	Friendly loss
11	92	Risk (expected loss, population risk, true risk, generalization error, expected prediction error)
12	93	Conditional risk (pointwise risk, risk at an input)
13	94	Empirical risk
14	95	Training error
15	96	Empirical risk minimization
16	97	Metric
17	98	Absolute metric
18	99	Relative metric
19	100	Accuracy
20	101	Error rate
21	102	Interpolation

Generalization and Cross-Validation

Table 7: The definitions on Generalization and Cross-Validation, in page order.

On page	Overall	Definition
1	103	Objective function (objective)
2	104	Cost function
3	105	Validation set
4	106	Validation error
5	107	Test set
6	108	Test error
7	109	Generalization gap
8	110	Folds
9	111	k-fold cross-validation
10	112	k-fold cross-validation estimate
11	113	Leave-one-out cross-validation

Bias, Variance and Expected Squared Error of a Prediction

Table 8: The definitions on Bias, Variance and Expected Squared Error of a Prediction, in page order.

On page	Overall	Definition
1	114	Prediction rule
2	115	Average fit
3	116	Bias of a prediction

On page	Overall	Definition
4	117	Variance of a prediction
5	118	Underfitting
6	119	Overfitting
7	120	Residual variance (residual noise variance, irreducible error)
8	121	Best member of a model family
9	122	Deterministic noise
10	123	Regularization
11	124	Penalty weight (regularization strength)

Kinds of Learning Problems and Features

Table 9: The definitions on Kinds of Learning Problems and Features, in page order.

On page	Overall	Definition
1	125	Sequence labeling
2	126	Sequence-to-sequence prediction (sequence-to-sequence decoding, seq2seq)
3	127	Ranking
4	128	Structured prediction
5	129	Density estimation
6	130	Anomaly detection (novelty detection, outlier detection)
7	131	Non-parametric model
8	132	Probabilistic model
9	133	Non-probabilistic model
10	134	Discriminative model
11	135	Generative model
12	136	One-hot encoding (indicator coding)
13	137	Standardization (z-scoring, standard scaling)
14	138	Min–max normalization (min–max scaling)

M4: Classification and Logistic Regression

Classification Problems

Table 10: The definitions on Classification Problems, in page order.

On page	Overall	Definition
1	139	Decision region
2	140	Decision boundary

Logistic Regression

Table 11: The definitions on Logistic Regression, in page order.

On page	Overall	Definition
1	141	Logistic regression
2	142	Link function
3	143	Probit regression
4	144	Complementary log-log regression

Gradient Descent

Table 12: The definitions on Gradient Descent, in page order.

On page	Overall	Definition
1	145	Learning rate (step size)
2	146	Stochastic gradient descent
3	147	Robbins–Monro conditions
4	148	Minibatch gradient descent
5	149	Epoch
6	150	Online training (online learning)
7	151	Patience
8	152	Performance decay

Fitting Logistic Regression

Table 13: The definitions on Fitting Logistic Regression, in page order.

On page	Overall	Definition
1	153	Log loss
2	154	Cross-entropy loss
3	155	Total log loss
4	156	Margin of an observation
5	157	Cross-entropy of a classifier
6	158	Residual of a logistic regression

Predictions and Inference for Logistic Regression

Table 14: The definitions on Predictions and Inference for Logistic Regression, in page order.

On page	Overall	Definition
1	159	Separated data (completely separated data)
2	160	Quasi-completely separated data
3	161	Hauck–Donner effect

Multiclass Logistic Regression

Table 15: The definitions on Multiclass Logistic Regression, in page order.

On page	Overall	Definition
1	162	Softmax function
2	163	Multinomial logistic regression (softmax regression)
3	164	Ordinal logistic regression (cumulative logit model, proportional odds model)

Discriminant Analysis

Table 16: The definitions on Discriminant Analysis, in page order.

On page	Overall	Definition
1	165	Bayes classifier
2	166	Generative classifier (plug-in Bayes classifier)
3	167	Linear discriminant function
4	168	Linear discriminant analysis (LDA)
5	169	Quadric surface, quadric
6	170	Quadratic discriminant function
7	171	Quadratic discriminant analysis (QDA)
8	172	The naive Bayes assumption
9	173	Naive Bayes classifier
10	174	Gaussian naive Bayes classifier
11	175	K-nearest neighbors classifier

M5: Performance Metrics

Metrics

Table 17: The definitions on Metrics, in page order.

On page	Overall	Definition
1	176	Confusion matrix for C classes
2	177	Confusion matrix for a detection task
3	178	Precision (also called positive predictive value)
4	179	Recall (also called sensitivity or true positive rate)
5	180	Specificity (also called true negative rate)
6	181	One-vs-rest reduction
7	182	Macro-averaging
8	183	Micro-averaging
9	184	Balanced accuracy
10	185	F-measure (also called F-score)
11	186	F1 score
12	187	Threshold

On page	Overall	Definition
13	188	Precision-recall curve
14	189	Area under the precision-recall curve (PR AUC)
15	190	False positive rate (also called false alarm rate)
16	191	Receiver operating characteristic (ROC) curve
17	192	Area under the ROC curve (ROC AUC)

M6: Regularization and Model Selection

Subset Selection

Table 18: The definitions on Subset Selection, in page order.

On page	Overall	Definition
1	193	Forward selection
2	194	Backward selection
3	195	Best subset selection

Regularization

Table 19: The definitions on Regularization, in page order.

On page	Overall	Definition
1	196	Ridge regression, Tikhonov regularization, ℓ_2 regularization
2	197	Weight decay
3	198	Lasso (least absolute shrinkage and selection operator), ℓ_1 regularization
4	199	Soft-thresholding operator
5	200	Cyclic coordinate descent
6	201	Derived feature
7	202	Dimension reduction regression
8	203	Principal components regression (PCR)
9	204	Partial least squares regression (PLS)
10	205	High-dimensional data set
11	206	Elastic net
12	207	Data augmentation

Model Selection

Table 20: The definitions on Model Selection, in page order.

On page	Overall	Definition
1	208	Hyperparameter, tuning parameter
2	209	Nested prediction function families, nested models
3	210	Variance of the cross-validation estimate
4	211	Selection bias, winner's curse

M7: Support Vector Machines and Kernel Methods

Support Vector Machines

Table 21: The definitions on Support Vector Machines, in page order.

On page	Overall	Definition
1	212	Linear classifier
2	213	Linearly separable
3	214	Separating hyperplane
4	215	Margin
5	216	Support vector
6	217	Max-margin hyperplane
7	218	Hard-margin primal problem
8	219	Hard-margin dual problem
9	220	Dual variable
10	221	Slack variable
11	222	C-SVM primal problem
12	223	C-SVM dual problem
13	224	Hinge loss
14	225	Kernel function
15	226	Kernel trick
16	227	One-versus-rest
17	228	One-versus-one

M8: Neural Networks and Backpropagation

Neural Networks

Table 22: The definitions on Neural Networks, in page order.

On page	Overall	Definition
1	229	Activation function
2	230	Rectified linear unit (ReLU)
3	231	Unit (neuron)
4	232	Pre-activation
5	233	Activation
6	234	Layer
7	235	Hidden layer
8	236	Neural network
9	237	Input layer

On page	Overall	Definition
10	238	Output layer
11	239	Feed-forward neural network
12	240	Autoencoder
13	241	Depth
14	242	Width
15	243	Deep learning
16	244	Fully connected layer
17	245	Dropout
18	246	Over-parametrized model (over-parameterized model)
19	247	Double descent

Backpropagation

Table 23: The definitions on Backpropagation, in page order.

On page	Overall	Definition
1	248	Backpropagation
2	249	Dynamic programming
3	250	Exploding gradient
4	251	He initialization
5	252	Glorot initialization
6	253	Gradient clipping
7	254	Saturated unit
8	255	Leaky ReLU (leaky rectifier)
9	256	Residual connection (skip connection, shortcut connection)
10	257	Saliency map

M9: ML Experimentation

Machine Learning Experimentation

Table 24: The definitions on Machine Learning Experimentation, in page order.

On page	Overall	Definition
1	258	Inner loop
2	259	Outer loop
3	260	Data snooping
4	261	Sampling bias
5	262	Early stopping

M10: Clustering

Unsupervised Learning

Table 25: The definitions on Unsupervised Learning, in page order.

On page	Overall	Definition
1	263	Proportion of variance explained (explained variance ratio)
2	264	Cumulative proportion of variance explained
3	265	Scree plot

Optional Reading: Theory

Machine Learning Theory

Table 26: The definitions on Machine Learning Theory, in page order.

On page	Overall	Definition
1	266	Inductive bias
2	267	Shattering
3	268	Growth function
4	269	VC dimension

Optional Reading: Non-Parametric Models

Non-Parametric Models

Table 27: The definitions on Non-Parametric Models, in page order.

On page	Overall	Definition
1	270	K nearest neighbors
2	271	K-nearest neighbors regression
3	272	Local linear regression
4	273	Span
5	274	LOESS

Splines

Table 28: The definitions on Splines, in page order.

On page	Overall	Definition
1	275	Polynomial regression
2	276	Step function regression, piecewise-constant regression
3	277	Piecewise polynomial
4	278	Cubic regression spline
5	279	Spline of degree d
6	280	Natural cubic spline
7	281	Smoothing spline
8	282	Generalized cross-validation and restricted maximum likelihood

Generalized Additive Models

Table 29: The definitions on Generalized Additive Models, in page order.

On page	Overall	Definition
1	283	Component function
2	284	Additive model
3	285	Generalized additive model
4	286	Backfitting

Optional Reading: Trees and Ensembles

Decision Trees

Table 30: The definitions on Decision Trees, in page order.

On page	Overall	Definition
1	287	Recursive binary splitting
2	288	Cost-complexity pruning, weakest-link pruning
3	289	Cross-validated cost-complexity pruning

Optional Reading: Further Methods

Gaussian Processes

Table 31: The definitions on Gaussian Processes, in page order.

On page	Overall	Definition
1	290	Gaussian process
2	291	Mean function
3	292	Covariance function (kernel)
4	293	Hyperparameters
5	294	Marginal likelihood

Convolutional and Recurrent Networks

Table 32: The definitions on Convolutional and Recurrent Networks, in page order.

On page	Overall	Definition
1	295	Flattening (vectorizing, unrolling)
2	296	Filter (kernel)
3	297	Convolution
4	298	Feature map (activation map, response map)
5	299	Convolutional layer
6	300	Detector layer
7	301	Convolutional neural network (CNN)
8	302	Stride

On page	Overall	Definition
9	303	Padding
10	304	Parameter sharing (weight sharing, weight tying)
11	305	Sparse connectivity
12	306	Locally connected layer
13	307	Pooling (subsampling)
14	308	Location invariance (translation invariance)
15	309	Channel
16	310	Recurrent layer
17	311	Hidden state (memory)
18	312	Recurrent neural network (RNN)
19	313	Unrolling (unfolding through time)
20	314	Backpropagation through time
21	315	Gate
22	316	Long short-term memory (LSTM)

Attention and Transformers

Table 33: The definitions on Attention and Transformers, in page order.

On page	Overall	Definition
1	317	Perplexity
2	318	Substitution
3	319	Insertion
4	320	Deletion
5	321	Minimum edit distance alignment
6	322	Edit distance (Levenshtein distance)
7	323	Token error rate