Cambridge University Press
978-1-108-48346-9 — Essentials of Pattern Recognition
Jianxin Wu
Index
More Information

Index

χ^2 kernel, *see* chi-squared kernel

accuracy, 46, 63
additive kernels, 168
alignment of data, 254
ANN, *see* approximate nearest neighbor
applications
  autonomous driving, 4
  face recognition, 5, **44**
  gait recognition, 5
  natural language processing, 5
  optical character recognition, 3
  pedestrian detection, 5
  speech recognition, 5
  speech synthesis, 5
  surveillance, 7
approximate nearest neighbor, 58, **61**
area under the precision–recall curve, 78
area under the ROC curve, 77
arithmetic mean, 204
AUC-PR, *see* area under the precision–recall curve
AUC-ROC, *see* area under the ROC curve
auto-conjugate, 305
average precision, 94

back propagation, 340
backward algorithm, HMM, 278
backward variable β, HMM, 278
basic problems in HMM, 274
  decoding, 274
  evaluation, 274
  parameter learning, 274
batch normalization, 362
batch processing, 339
Baum–Welch algorithm, 283, **287**
  EM interpretation, 331
Bayes decision theory, 73
Bayes error rate, 80
Bayes' theorem, 28, 175
Bayesian estimation, 178, 181
beta distribution, 314
beta function, 314
bias–variance decomposition, 83
  bias, 83
  variance, 83
BN, *see* batch normalization

canonical parameterization, Gaussian, 296
categorical data, 56
categorical variable, *see* nominal variable
Cauchy distribution, 30
Cauchy–Schwarz inequality, 17
  integral form, 17
  Schwarz inequality, 41
chain rule, 335
Chapman–Kolmogorov equations, 290
characteristic function, 298, **309**
  normal distribution, 309
Chebyshev's inequality, 31
  one-sided version, 31
chi-squared kernel, 169
Cholesky factorization, 139
city block distance, *see* Manhattan distance
class conditional distribution, 175
classification, 56, **174**
clustering, 57
CNN, *see* convolutional neural network
complete data log-likelihood, 322
compressive sensing, *see* sparse coding
computer vision, 5
concave function, 36
concept drift, 55
condition number, 137
  2-norm, 137
conditional entropy, 222
conditional independence, 277, **287**
conditionally positive definite kernel, 170
confusion matrix, 96
  normalized confusion matrix, 97
conjugate prior, 183
convex function, 36
  strictly convex, 37
convex set, 36
convolution, 344

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-48346-9 — Essentials of Pattern Recognition
Jianxin Wu
Index
More Information

convolution kernel, 344
convolution stride, 346
convolutional neural network, 333
  backward run, 336
  forward run, 336, **337**
cost matrix, 72
cost minimization, *see* loss minimization
cost sensitive, 72
cross entropy, 230
cross entropy loss, 232
curse of dimensionality, 187

δ variables, HMM, 282
d-separation rule, 289
DAG, *see* directed acyclic graph
decision forest, 237
decision tree, 235
  attribute, 235
deep learning, 3, 348
  end-to-end, 10
dictionary, 251
  overcomplete, 251
  undercomplete, 251
dictionary learning, 251
differential entropy, 226
dimensionality reduction, 103
  linear dimensionality reduction, 104
directed acyclic graph, 363
discrete metric, 197
discrete-time Markov chain, *see* Markov chain
discriminant function, 176
discriminative model, 176
distance metric learning, 202
distributed representation, 348
divide and conquer, 258
DTMC, *see* Markov chain
DTW, *see* dynamic time warping
dual coordinate descent, 171
duality gap, 154
dynamic programming, 258
dynamic time warping, 255–262

early stopping, 68
eigendecomposition, 23
EM, *see* expectation-maximization
empirical loss, 159
entropy, 221
epoch, 339
error rate, 63
  generalization error, 63
  test error, 64
  training error, 64
  validation error, 68
Euclidean distance, 197
expectation-maximization, 327
explaining away, 289

exponential distribution, 42, 192
  entropy, 227
  memoryless, 42
exponential family, 312
  conjugate prior, 315

F_β measure, 79
F-measure, 78
face recognition
  PCA+FLD, 124, 139
false negative, 16, 74
false negative rate, 75
false positive, 16, 74
false positive rate, 75
fast iterative shrinkage-thresholding algorithm, 264
feature extraction, 7
feature mapping, 164
  feature space, 164
  input space, 164
feature selection, 233
  mRMR, 234
  sparse linear classifier, 252
features, 9
  feature learning by deep learning, 9
  feature vector, 9
  manually designed, 9
Fisher's linear discriminant, 125–132
  between-class scatter matrix, 129
  multiclass extension, 132
  total scatter matrix, 134
  within-class scatter matrix, 129
FISTA, *see* fast iterative shrinkage-thresholding algorithm
FLD, *see* Fisher's linear discriminant
forward algorithm, HMM, 277
forward variable α, HMM, 277
Frobenius norm, 121

γ variables, HMM, 280
  unnormalized, 281
Gauss transformation, 138
Gaussian distribution, *see* normal distribution
Gaussian kernel, *see* kernel function, radial basis function kernel
Gaussian mixture model, 177, **317**
  EM algorithm, 328–330
  mixing coefficients, 317
generalized mean, 169, **203**
generalized Rayleigh quotient, 130
generating a sequence
  in DTMC, 272
  in HMM, 274
generative model, 176
geometric distribution (discrete), 26
  entropy, 225
geometric mean, 204

Gibbs' inequality, 241
Givens rotation, 120
GMM, *see* Gaussian mixture model
gradient, 338
gradient descent, 338
graphical model, *see* probabilistic graphical model
groundtruth labels, 49, **80–81**
group sparsity, 250

harmonic mean, 78, 204
Hellinger's kernel, 169
Hessian, 37
hidden Markov model, 272
hidden variables, 174, **319**
hinge loss, 159
histogram, 185
  bin, 185
  bin width, 185
histogram intersection kernel, 168
HMM, *see* hidden Markov model
Huffman code, 220
Huffman tree, 220
hyperbolic tangent function, 212
hyperparameter, 65, 165

i.i.d., *see* independent and identically distributed
ill-conditioned function, 137
ill-conditioned matrix, 137
imbalanced classification, 52
  evaluation, 72
  SVM learning, 168
incomplete data log-likelihood, 322
independent and identically distributed, 55
indicator function, 64
information gain, 238
information theory, 57, 221
inherent dimensionality, 101
integral image, 265
ISTA, iterative shrinkage-thresholding algorithms, 263

Jacobi method, 121
Jensen's inequality, 36
joint entropy, 222

k-fold cross-validation, 69
K-means clustering, 62
  Lloyd's method, 62
Kalman filter, 269, 305
  innovation, 307
  Kalman gain matrix, 307
Karush–Kuhn–Tucker conditions, 153
KDE, kernel density estimation, 188
kernel density estimation
  bandwidth, 189
  theoretically optimal bandwidth, 189
kernel function in KDE, 188
  Epanechnikov kernel, 189
  Gaussian kernel, 189
kernel functions, 165
  additive kernels, 168
  linear kernel, 165
  polynomial kernel, 165
  power mean kernels, 169
  radial basis function kernel, 165
kernel matrix, 163
kernel methods, 161
kernel trick, 164
KKT conditions, *see* Karush–Kuhn–Tucker conditions
KL distance, *see* Kullback–Leibler divergence
Kronecker product, 350
Kullback–Leibler divergence, 223

ℓ_1 norm
  induce sparsity, 246
ℓ_2 normalization, 210
ℓ_0 "norm", 199
ℓ_1 distance, 199
ℓ_1 norm, 199
ℓ_1 normalization, 210
ℓ_p distance, 199
Laplace distribution, 227
  entropy, 227
large margin classifiers, 147
lasso, 252
layers, CNN, 336
  ReLU layer, 342
  softmax layer, 337
  average pooling layer, 357
  convolution layer, 344
  dropout layer, 360, **362**
  fully connected layer, 356
  loss layer, 337
  max pooling layer, 357
LDA, *see* linear discriminant analysis
LDL factorization, 138
learning, 45
  training set, 45
learning rate, 338
likelihood, 175
linear discriminant analysis, 131
linear model, 9
linear regression, 93, **207**
linearly separable, 144
Lipschitz constant, 263
LLE, *see* locally linear embedding
locally linear embedding, 213
log partition function, 313
log sum inequality, 241
log-likelihood function, 180
log-normal distribution, 193
  entropy, 227

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-48346-9 — Essentials of Pattern Recognition
Jianxin Wu
Index
More Information

logistic regression, 212
logistic sigmoid function, 211
loss minimization, 71
LU factorization, 138

Mahalanobis distance, 202, **300**
Manhattan distance, 199
MAP, *see* maximum a posteriori estimation
margin, 147
  of a dataset, 147
  of one example, 147
margin calculation, 147–149
  formula, 149
  geometry, 148
marginal likelihood, 174
Markov chain, 270
Markov property, 269
  in HMM, 273
Markov's inequality, 31
matrix inversion lemma, 311
matrix norm, 137
  matrix 2-norm, 137
matrix rank, 136
  column rank, 136
  full rank, 136
  row rank, 136
max-margin classifiers, 147
maximum a posteriori estimation, 181
maximum entropy distribution
  exponential distribution, 241
  multivariate normal, 229
maximum likelihood estimation, 178, **180**
  likelihood, 179
  likelihood function, 179
mean field approximation, 193
mean squared error, 71
Mercer's condition, 163
message passing, 283
method of Lagrange multipliers, 38, 152
  Lagrange function, 38
  Lagrange multiplier, SVM, 152
  Lagrange multipliers, 38
metric, 197
  induced by vector norm, 198
mini-batch, 340
Minkowski inequality, 199
mixed $\ell_{2,1}$ (matrix) norm, 249
mixed $\ell_{\alpha,\beta}$ matrix norm, 249
MLE, *see* maximum likelihood estimation
model, 9
  model learning, 9
  parameters, 9
model selection, 274
moment parameterization, Gaussian, 296
Moore–Penrose pseudoinverse, 132
MSE, *see* mean squared error
multinomial distribution, 318
multinomial logistic regression, 232
multivariate KDE, 190
  diagonal bandwidth matrix, 190
multivariate normal distribution, 32, **294**
  entropy, 229
mutual information, 223

nat, 226
nearest neighbor classifier, 45
  k-nearest neighbors, 48
no free lunch theorem, 51
noise, 83
nominal variable, 174
nonparametric methods, 177
normal distribution, 32, **293**
  entropy, 226

observation probability matrix, 273
observations, 272
observed variables, 173, **319**
OCR, *see* applications, optical character recognition
one-vs.-one, 160
one-vs.-rest, 160
ordinary linear regression, 93
outlier, 55, **103**
overfitting, 66, 187

ψ variables, HMM, 282
p-norm, 198
pad an image, 345
parameter estimation, 176
parametric methods, 177
Pareto distribution, 192
  conjugate prior, 193
partition function, 313
PCA, *see* principal component analysis
Pearson's correlation coefficient, 31
per-dimension normalization, 209
per-example normalization, 210
point estimation, 178
polynomial regression, 65, **94**
  degree, 65
posterior distribution, 175
posterior predicative distribution, 184
power mean kernels, 169, **204**
PR curve, *see* precision–recall curve
precision, 77
precision–recall curve, 78
prefix code, 220
principal component analysis, 111
  principal component, 111
  projection direction, 109

© in this web service Cambridge University Press www.cambridge.org

Cambridge University Press
978-1-108-48346-9 — Essentials of Pattern Recognition
Jianxin Wu
Index
More Information

rule of thumb for *d*, 114
zero-dimensional representation, 105
principle of maximum entropy, 231
prior distribution, 174
probabilistic graphical model, 271
directed, 273, **288**
undirected, 271
probabilistic inference, 175
Pythagorean theorem, 149

QR decomposition, 121

random process, *see* stochastic process
Rayleigh quotient, 42
RBF kernel, *see* kernel function, radial basis function kernel
recall, 77
receiver operating characteristics curve, 76
receptive field, 361
rectified linear unit, 342
recurrent neural network, 267
nonlinear activation function, 268
output vector, 268
state vector, 268
unfold, 268
regression, 56, **174**
dependent variable, 174
independent variables, 174
regularization term, 71
regularizer, *see* regularization term, 159
relative entropy, *see* Kullback–Leibler divergence
ReLU, *see* rectified linear unit
representation learning, *see* deep learning
representer theorem, 153, 164
ResNet, 363
ridge regression, 93, **216**
RNN, *see* recurrent neural network
ROC curve, *see* receiver operating characteristics curve

saturated function, 343
scatter matrix, 128
Schur complement, 301, **310**
self-information, *see* entropy
semantic gap, 50
sensors, 6
sequential data, 254
sequential minimal optimization, 166
SGD, *see* stochastic gradient descent
singular value decomposition, 24
slack variable, 158
SMO, *see* sequential minimal optimization
Sobel operator, 346
soft thresholding, 247, 262
softmax regression, *see* multinomial logistic regression
softmax transform, 211
sparse coding, 250
sparse logistic regression, 253
sparse support vector classification, 253
sparse vector, 245
spectral decomposition, **23**, 110
spherical normal distribution, 295
square integrable function, 163
state transition probability matrix, 271
n-step transition matrix, 290
stochastic gradient descent, 338, **340**
stochastic matrix, 271
left stochastic matrix, 271
right stochastic matrix, 271
stochastic process, 269
stratified sampling, 96
string matching, 255
structured sparsity, 250
Student's *t*-distribution, 87, 90–91
degrees of freedom, 90
Student's *t*-test, 88–92
null hypothesis, 89
one-tailed, 92
paired *t*-test, 88
significance level, 89
test statistics, 89
two-tailed, 91
submodular function, 234
subspace methods, 104
sufficient statistics, 312
summed area table, *see* integral image
supervised learning, 57
support vector machines, 155
complementary slackness, **154**, 155
dual form, linear, inseparable, 159
dual form, linear, separable, 154
dual form, nonlinear, 163
multiclass, 160
primal form, linear, inseparable, 159
primal form, linear, separable, 152
sparsity, 157, 249
without bias, 170
support vector regression, 166
support vectors, 155
surrogate loss function, 71, 246
SVD, *see* singular value decomposition
SVM, *see* support vector machines
SVR, *see* support vector regression

testing, 46
test set, 46
time window, 266
training example, 45
label, 45

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-108-48346-9 — Essentials of Pattern Recognition
Jianxin Wu
Index
More Information

true negative, 16, 74
true negative rate, 75
true positive, 16, 74
true positive rate, 75

underfitting, 65, 187
universal approximator, 177
unsupervised learning, 57

validation set, 67
vector norm, 198
Viterbi decoding algorithm, 283

whitening transform, 118

ξ variables, HMM, 284

© in this web service Cambridge University Press www.cambridge.org