

Index

Symbols

L_2 norm, 321
 χ^2 -distribution, 171
 χ^2 -statistic, 171
3D bar chart, 30

A

absorbing state, 332
accuracy, 254
affinity, 290
Agglomerative Clustering, 283
all-vs-all, 268
analysis of variance, 182
ANOVA, 182
ANOVA table, 182
approximate nearest neighbor, 256
Approximating a symmetric matrix with a low rank matrix, 357
Approximating a symmetric matrix with low dimensional factors, 358
average, 7

B

bag, 272
bagging, 272
balanced, 180
balanced experiment, 183
bar chart, 5
baselines, 254
Basic properties of the probability events, 55
batch, 263
batch size, 263
Bayes risk, 254
Bayes' rule, 89
Bayesian inference, 207
Bayesian inference is particularly good with little data, 211
Bernoulli random variable, 116
Beta distribution, 120
between group variation, 182
biased estimate, 147
biased random walk, 331
bigram models, 337
bigrams, 337
bimodal, 16
Binomial distribution, 117
Binomial distribution for large N , 130
bootstrap, 152
bootstrap replicates, 152
box plot, 20
Building a decision forest, 272
Building a decision forest using bagging, 273
Building a decision tree: overall, 271

C

categorical, 3
Centered confidence interval for a population mean, 146
Chebyshev's inequality, 100
class conditional probability, 257
class confusion matrix, 254
class error rate, 254
class-conditional histograms, 7
Classification with a decision forest, 273
Classifier, 253
classifier, 253
 definition, 253
 nearest neighbors, 256
cluster center, 283
clustering, 281
 using K-means, 287
clusters, 283
color constancy, 241
comparing to chance, 254
complete-link clustering, 283
Computing a one-sided p-value for a T-test, 163
Computing a two-sided p-value for a T-test, 163
conditional histograms, 7
Conditional independence, 72
Conditional probability, 66
Conditional probability for independent events, 71
Conditional probability formulas, 70
Confidence interval for a population mean, 146
conjugacy, 209
conjugate prior, 209
consistency, 206
Constructing a centered $1 - 2\alpha$ confidence interval for a population mean for a large sample, 151
Constructing a centered $1 - 2\alpha$ confidence interval for a population mean for a small sample, 151
continuous, 3
contrasts, 185
correlation, 36, 39
Correlation coefficient, 39
cost to go function, 345
Covariance, 97, 227
covariance ellipses, 302
Covariance Matrix, 229
Covariance, useful expression, 97
cross-validation, 255
Cumulative distribution of a discrete random variable, 88

D

decision function, 269
decision boundary, 260

decision forest, 269
 decision tree, 105, 268
 degrees of freedom, 147
 dendrogram, 284
 density, 91
 dependent variable, 305
 descent direction, 263
 descriptive statistics, 98
 diagonal, 355
 diagonalizing, 356
 Diagonalizing a symmetric matrix, 233
 Discrete random variable, 87
 distributions

- how often a normal random variable is how far from the mean, 125
- mean and variance of a bernoulli random variable, 116
- mean and variance of a beta distribution, 121
- mean and variance of a geometric distribution, 116
- mean and variance of the binomial distribution, 117
- mean and variance of the exponential distribution, 122
- mean and variance of the gamma distribution, 121
- mean and variance of the normal distribution, 124
- mean and variance of the poisson distribution, 119
- mean and variance of the standard normal distribution, 123

 Divisive Clustering, 284
 dynamic programming, 345

E

Easy confidence intervals for a big sample, 150
 eigenvalue, 232, 355
 eigenvector, 232, 355
 emission distribution, 344
 empirical distribution, 99, 152
 entropy, 360
 epoch, 264
 error, 254
 error bars, 150
 error function, 125, 358
 Estimating Confidence Intervals for Maximum Likelihood Estimates using Simulation, 205
 Estimating with maximum likelihood, 199
 Evaluating whether a treatment has significant effects with a one-way ANOVA for balanced experiments, 183
 Event, 55
 Expectation, 94
 Expectation of a continuous random variable, 95
 Expectations are linear, 95
 Expected value, 93
 Expected value of a continuous random variable, 95
 explanatory variables, 305
 Exponential distribution, 122
 Expressions for mean and variance of the sample mean, 144

F

F-distribution, 170
 F-statistic, 170
 false positive rate, 254
 false negative rate, 254
 feature vector, 253
 filtering, 214
 fold, 255
 Forming and interpreting a two-way ANOVA table, 191
 Frobenius norm, 356

G

gambler's fallacy, 64
 Gamma distribution, 121
 gaussian distributions, 124
 generalizing badly, 255
 Geometric distribution, 116
 gradient descent, 263
 group average clustering, 283

H

heat map, 30
 hidden Markov model, 344
 hinge loss, 261
 histogram, 6

I

IID, 198
 iid samples, 99
 independent and identically distributed, 198
 Independent events, 62
 independent identically distributed samples, 99
 Independent random variables, 90
 Independent random variables have zero covariance, 97
 indicator function, 100
 Indicator functions, 100
 information gain, 270, 361
 intensity, 119
 interaction mean squares, 189
 Interquartile Range, 15
 inverse error function, 358
 irreducible, 335

J

joint, 210
 Joint probability distribution of two discrete random variables, 89

K

k-means, *see* clustering, 287
 K-Means Clustering, 287
 K-Means with Soft Weights, 291

L

latent variable, 45
 learning curves, 264
 learning rate, 264
 leave-one-out cross-validation, 255
 Likelihood, 198
 likelihood, 257
 Likert scales, 247
 line search, 263
 Linear regression, 308
 Linear Regression using Least Squares, 312
 location parameter, 9
 Log-likelihood of a dataset under a model, 201

M

Many Markov chains have stationary distributions, 336
 MAP estimate, 207
 Marginal probability of a random variable, 90

- Markov chain, 331
- Markov chains, 333
- Markov's inequality, 100
- maximum a posteriori estimate, 207
- Maximum likelihood principle, 198
- Mean, 7
- mean and variance of
 - a bernoulli random variable, 116
 - a beta distribution, 121
 - a geometric distribution, 116
 - the binomial distribution, 117
 - the exponential distribution, 122
 - the gamma distribution, 121
 - the normal distribution, 124
 - the poisson distribution, 119
 - the standard normal distribution, 123
- Mean or expected value, 96
- mean square error, 310
- Median, 13
- mode, 16
- multidimensional scaling, 245
- multimodal, 16
- Multinomial distribution, 118

- N**
- n-gram models, 337
- n-grams, 337
- normal distribution, 124
- Normal data, 19
- Normal distribution, 124
- normal distribution, 124
- Normal posteriors can be updated online, 215
- normal quantile function, 358
- normal random variable, 124
- normalizing, 92
- normalizing constant, 92

- O**
- odds, 104
- one factor, 182
- one-sided p-value, 163
- one-vs-all, 268
- ordinal, 3
- orthonormal, 355
- Orthonormal matrices are rotations, 233
- outcomes, 53
- outlier, 13
- outliers, 314
- overfitting, 255

- P**
- p-value, 162
- p-value hacking, 174
- Pairwise independence, 72
- Parameters of a Multivariate Normal Distribution, 301
- pdf, 91
- Percentile, 14
- phonemes, 344
- pie chart, 29
- Poisson distribution, 119
- Poisson point process, 119
- population, 141
- population mean, 141
- positive definite, 355
- positive semidefinite, 355
- posterior, 207, 258
- Predicting a value using correlation, 43
- Predicting a value using correlation: Rule of thumb - 1, 44
- Predicting a value using correlation: Rule of thumb - 2, 44
- principal components, 237
- Principal Components Analysis, 240
- Principal Coordinate Analysis, 246
- principal coordinate analysis, 245
- prior, 257
- prior probability distribution, 207
- probability, 54
- probability density function, 91
- Probability distribution of a discrete random variable, 88
- probability mass function, 88
- probit function, 358
- procedure
 - predicting a value using correlation, 43
 - the t-test of significance for a hypothesized mean, 162
 - agglomerative clustering, 283
 - approximating a symmetric matrix with a low rank matrix, 357
 - approximating a symmetric matrix with low dimensional factors, 358
 - building a decision forest, 272
 - building a decision forest using bagging, 273
 - building a decision tree: overall, 271
 - classification with a decision forest, 273
 - computing a one-sided p-value for a t-test, 163
 - computing a two-sided p-value for a t-test, 163
 - constructing a centered $1 - 2\alpha$ confidence interval for a population mean for a large sample, 151
 - constructing a centered $1 - 2\alpha$ confidence interval for a population mean for a small sample, 151
 - diagonalizing a symmetric matrix, 233
 - divisive clustering, 284
 - estimating confidence intervals for maximum likelihood estimates using simulation, 205
 - estimating with maximum likelihood, 199
 - evaluating whether a treatment has significant effects with a one-way anova for balanced experiments, 183
 - forming and interpreting a two-way anova table, 191
 - k-means clustering, 287
 - k-means with soft weights, 291
 - linear regression using least squares, 312
 - predicting a value using correlation: rule of thumb—1, 44
 - predicting a value using correlation: rule of thumb—2, 44
 - principal components analysis, 240
 - principal coordinate analysis, 246
 - setting up a two-way anova, 191
 - splitting a non-ordinal feature, 272
 - splitting an ordinal feature, 272
 - testing whether two populations have the same mean, for different population standard deviations, 169
 - testing whether two populations have the same mean, for known population standard deviations, 166
 - testing whether two populations have the same mean, for same but unknown population standard deviations, 167
 - the χ^2 -test of significance of fit to a model, 172
 - the bootstrap, 153
 - the f-test of significance for equality of variance, 170
 - training an svm: estimating the accuracy, 266
 - training an svm: overall, 266
 - training an svm: stochastic gradient descent, 267
 - vector quantization—building a dictionary, 296
 - vector quantization—representing a signal, 296

Properties of normal data, 20
 Properties of probability density functions, 92
 Properties of sample and population means, 142
 Properties of standard deviation, 10
 Properties of the correlation coefficient, 40
 Properties of the covariance matrix, 230
 Properties of the interquartile range, 15
 Properties of the mean, 8
 Properties of the median, 14
 Properties of the probability of events, 59
 Properties of variance, 13, 96
 prosecutor's fallacy, 72

Q

Quartiles, 14

R

Randomization, 179
 raw Google matrix, 343
 realization, 99
 recurrent, 332
 Regression, 305, 308, 311
 regularization, 262
 regularization parameter, 262
 regularization weight, 319
 regularizer, 262
 residual, 310
 residual variation, 181
 ridge regression, 319

S

sample, 99, 141
 sample mean, 141
 Sample space, 53
 scale parameter, 10
 scatter plot, 33
 selection bias, 255
 sensitivity, 254
 Setting up a two-way ANOVA, 191
 single-link clustering, 283
 skew, 16
 smoothing, 338
 specificity, 254
 Splitting a non-ordinal feature, 272
 Splitting an ordinal feature, 272
 stacked bar chart, 30
 Standard coordinates, 18
 Standard deviation, 9, 98
 standard deviation, 98
 Standard error, 147
 standard normal curve, 19
 Standard normal data, 19
 Standard Normal distribution, 123
 standard normal distribution, 123
 standard normal random variable, 123
 stationary distribution, 335
 statistic, 146
 Statistical significance, 162
 step size, 264
 steplength, 264
 steplength schedule, 264
 Stochastic gradient descent, 263
 stochastic matrices, 333

Sums and differences of normal random variables, 165
 support vector machine, 261
 SVM, 261
 symmetric, 232, 355

T

T-distribution, 149
 T-random variable, 149
 T-test, 162
 tails, 16
 test error, 255
 test examples, 305
 test statistic, 160
 Testing whether two populations have the same mean, for different population standard deviations, 169
 Testing whether two populations have the same mean, for known population standard deviations, 166
 Testing whether two populations have the same mean, for same but unknown population standard deviations, 167
 The χ^2 -test of significance of fit to a model, 172
 The bootstrap, 153
 The F-test of significance for equality of variance, 170
 The parameters of a normal posterior with a single measurement, 214
 The properties of simulations, 341
 The T-test of significance for a hypothesized mean, 162
 total error rate, 254
 Training an SVM: estimating the accuracy, 266
 Training an SVM: Overall, 266
 Training an SVM: stochastic gradient descent, 267
 training error, 255
 training examples, 305
 transition probabilities, 331
 Transition probability matrices, 335
 treatment one mean squares, 190
 treatment two mean squares, 190
 treatment variation, 182
 trellis, 345
 trial, 99
 trigram models, 337
 trigrams, 337
 two-factor ANOVA, 191
 two-sided p-value, 163
 two-way ANOVA, 191

U

unbalanced experiment, 183
 unbiased, 255
 unbiased estimate, 147
 uniform distribution, 120
 Uniform distribution, continuous, 120
 uniform random variable, 120
 Uniform random variable, discrete, 115
 unigram models, 337
 unigrams, 337
 unimodal, 16
 useful facts
 basic properties of the probability events, 55
 bayesian inference is particularly good with little data, 211
 binomial distribution for large n , 130
 conditional probability for independent events, 71
 conditional probability formulas, 70
 covariance, useful expression, 97
 easy confidence intervals for a big sample, 150
 expectations are linear, 95

expressions for mean and variance of the sample mean, 144
how often a normal random variable is how far from the mean, 125
independent random variables have zero covariance, 97
many markov chains have stationary distributions, 336
markov chains, 333
mean and variance of a bernoulli random variable, 116
mean and variance of a beta distribution, 121
mean and variance of a geometric distribution, 116
mean and variance of the binomial distribution, 117
mean and variance of the exponential distribution, 122
mean and variance of the gamma distribution, 121
mean and variance of the normal distribution, 124
mean and variance of the poisson distribution, 119
mean and variance of the standard normal distribution, 123
normal posteriors can be updated online, 215
orthonormal matrices are rotations, 233
parameters of a multivariate normal distribution, 301
properties of normal data, 20
properties of probability density functions, 92
properties of sample and population means, 142
properties of standard deviation, 10
properties of the correlation coefficient, 40
properties of the covariance matrix, 230
properties of the interquartile range, 15
properties of the median, 14
properties of the probability of events, 59
properties of variance, 13, 96
regression, 311
sums and differences of normal random variables, 165
the parameters of a normal posterior with a single measurement,
214
the properties of simulations, 341

transition probability matrices, 335
variance as covariance, 98
variance, a useful expression, 96
you can transform data to zero mean and diagonal covariance, 234
utility, 107

V

validation set, 255
Variance, 13, 96
Variance as covariance, 98
Variance, a useful expression, 96
vector quantization, 296
Vector Quantization - Building a Dictionary, 296
Vector Quantization - Representing a Signal, 296
Viterbi algorithm, 345

W

Weak Law of Large Numbers, 102
whitening, 256, 286
within group variation, 181
within group mean squares, 189

Y

You can transform data to zero mean and diagonal covariance, 234

Z

Zipf's law, 313