Index

accuracy, 415
activation, 400
activation function, 401
additive, 11, 94–98, 110–111
additivity, 305, 306
adjusted R^2, 87, 231, 232, 236–238
`Advertising` data set, 15, 16, 19, 69, 71–73, 77, 78, 80, 82, 83, 85, 87–90, 95, 96, 109–111
agglomerative clustering, 525
Akaike information criterion, 87, 231, 232, 236–238
alternative hypothesis, 76, 559
analysis of variance, 312
ANOVA, 587
area under the curve, 155, 486–487
argument, 40
array, 42
attribute, 42
AUC, 155
`Auto` data set, 12, 66, 98–101, 129, 197, 202–207, 327, 398
auto-correlation, 421
autoregression, 423
axes, 48

backfitting, 307, 328
backpropagation, 429
backward stepwise selection, 87, 234–235
bag-of-n-grams, 415
bag-of-words, 414
bagging, 11, 24, 331, 343–346, 354, 360–361
BART, 343, 350, 353, 354, 362–363
baseline, 93, 145, 161
basis function, 293–294, 296
Bayes
  classifier, 35–37, 147
  decision boundary, 148
  error, 35–37
Bayes' theorem, 146, 250
Bayesian, 250–251, 353
Bayesian additive regression trees, 331, 343, 350, 350, 353, 354, 362–363
Bayesian information criterion, 87, 231, 232, 236–238
Benjamini–Hochberg procedure, 575–577
Bernoulli distribution, 172
best subset selection, 231, 246
bias, 31–34, 74, 90, 159, 405
bias-variance
  decomposition, 32

© Springer Nature Switzerland AG 2023

G. James et al., *An Introduction to Statistical Learning*, Springer Texts in Statistics,
https://doi.org/10.1007/978-3-031-38747-0

trade-off, 31–34, 38, 111–112, 157, 159, 163, 164, 242, 254, 263, 266, 301, 336, 376, 385
bidirectional, 425
Bikeshare data set, 12, 167–172
binary, 27, 138
biplot, 507, 508
Bonferroni method, 575–577, 585
Boolean, 53, 176
boosting, 11, 24, 331, 343, 347–350, 354, 361–362
bootstrap, 11, 201, 212–214, 343
Boston data set, 12, 67, 117, 122, 133, 199, 227, 287, 327, 364, 556
bottom-up clustering, 525
boxplot, 62
BrainCancer data set, 12, 472–474, 476, 482
branch, 333
burn-in, 352

C-index, 487
Caravan data set, 12, 184, 366
Carseats data set, 12, 126, 130, 364
categorical, 2, 27
censored data, 469–502
censoring
independent, 471
interval, 471
left, 471
mechanism, 471
non-informative, 471
right, 471
time, 470
chain rule, 429
channel, 407
CIFAR100 data set, 406, 409–411, 448, 449
classification, 2, 11, 27, 34–39, 135–199, 367–382
error rate, 338
tree, 337–341, 355–358
classifier, 135
cluster analysis, 25–26
clustering, 4, 25–26, 520–535
agglomerative, 525
bottom-up, 525
hierarchical, 521, 525–535
K-means, 11, 521–524
Cochran–Mantel–Haenszel test, 475
coefficient, 71
College data set, 12, 65, 286, 328
collinearity, 106–110
concatenation, 41
conditional probability, 35
confidence interval, 75–76, 90, 110, 292
confounding, 144
confusion matrix, 153, 176
continuous, 2
contour, 246
contour plot, 50
contrast, 94
convenience function, 53
convolution filter, 407
convolution layer, 407
convolutional neural network, 406–413
correlation, 79, 82–83, 530
count data, 167, 170
Cox's proportional hazards model, 480, 483–486
C_p, 87, 231, 232, 236–238
Credit data set, 12, 91, 92, 94, 97, 98, 106–109
cross-entropy, 405
cross-validation, 11, 31, 34, 201–211, 231, 252, 270
k-fold, 206–209
leave-one-out, 204–206
curse of dimensionality, 115, 193, 266

data augmentation, 411
data frame, 55
Data sets
Advertising, 15, 16, 19, 69, 71–73, 77, 78, 80, 82, 83, 85, 87–90, 95, 96, 109–111
Auto, 12, 66, 98–101, 129, 197, 202–207, 327, 398
Bikeshare, 12, 167–172

Boston, 12, 67, 117, 122, 133, 199, 227, 287, 327, 364, 556
BrainCancer, 12, 472–474, 476, 482
Caravan, 12, 184, 366
Carseats, 12, 126, 130, 364
CIFAR100, 406, 409–411, 448, 449
College, 12, 65, 286, 328
Credit, 12, 91, 92, 94, 97, 98, 106–109
Default, 12, 136–139, 141–144, 152–156, 160, 161, 225, 226, 466
Fund, 12, 567–570, 572, 575, 576, 585, 588, 589
Heart, 339, 340, 344–347, 352, 353, 382, 383
Hitters, 12, 332, 333, 336, 338, 339, 366, 425, 426, 437, 446
IMDb, 413, 415, 416, 418, 420, 437, 458, 467
Income, 16–18, 21–23
Khan, 12, 579–581, 583, 590, 593
MNIST, 402–404, 406, 430, 431, 441, 444, 445, 448
NCI60, 4, 5, 12, 546, 548–550
NYSE, 12, 422–424, 466, 467
OJ, 12, 365, 398
Portfolio, 12
Publication, 12, 482–487
Smarket, 2, 3, 12, 173, 184, 196
USArrests, 12, 507, 508, 510, 512, 513, 515, 516, 518, 519
Wage, 1, 2, 8, 9, 12, 290, 291, 293, 295, 297–300, 302–306, 309, 315, 327
Weekly, 12, 196, 226
data type, 42
decision function, 387
decision tree, 11, 331–342
deep learning, 399
Default data set, 12, 136–139, 141–144, 152–156, 160, 161, 225, 226, 466
degrees of freedom, 30, 266, 295, 296, 301
dendrogram, 521, 525–530
density function, 146
dependent variable, 15
derivative, 296, 300
detector layer, 410
deviance, 232
dictionary, 66
dimension reduction, 230, 253–262
discriminant function, 149
discriminant method, 146–161
dissimilarity, 530–532
distance
  correlation-based, 530–532, 554
  Euclidean, 509, 522, 523, 529–532
double descent, 431–435
double-exponential distribution, 251
dropout, 406, 431
dummy variable, 91–94, 138, 142, 292

early stopping, 430
effective degrees of freedom, 301
eigen decomposition, 506, 516
elbow, 548
embedding, 418
embedding layer, 419
ensemble, 343–354
entropy, 337–339, 363
epochs, 430
error
  irreducible, 17, 30
  rate, 34
  reducible, 17
  term, 16
Euclidean distance, 509, 522, 523, 529–532, 554
event time, 470
exception, 45
expected value, 18
exploratory data analysis, 504
exponential, 173
exponential family, 173

F-statistic, 84

factor, 92
factorial, 170
failure time, 470
false
  discovery proportion, 155, 573
  discovery rate, 558, 573–577, 579–582
  negative, 155, 562
  positive, 155, 562, 563
  positive rate, 155, 156, 382
family-wise error rate, 565–573, 577
feature, 15
feature map, 406
feature selection, 230
featurize, 414
feed-forward neural network, 400
figure, 48
fit, 21
fitted value, 101
flattening, 424
flexible, 21
floating point, 43
forward stepwise selection, 86, 87, 233–234, 268
function, 40
Fund data set, 12, 567–570, 572, 575, 576, 585, 588, 589

Gamma, 173
Gaussian (normal) distribution, 146, 147, 150, 172, 561
generalized additive model, 5, 24, 162, 289, 290, 305–309, 319
generalized linear model, 5, 135, 167–174, 217
generative model, 146–161
Gini index, 337–339, 345, 346, 363
global minimum, 427
gradient, 428
gradient descent, 427

Harrell's concordance index, 487
hazard function, 476–478
  baseline, 478
hazard rate, 476
Heart data set, 339, 340, 344–347, 352, 353, 382, 383
heatmap, 50
helper, 311
heteroscedasticity, 103, 168
hidden layer, 400
hidden units, 400
hierarchical clustering, 525–530
  dendrogram, 525–528
  inversion, 529
  linkage, 529–530
hierarchical principle, 96
high-dimensional, 86, 234, 263
hinge loss, 385
Hitters data set, 12, 332, 333, 336, 338, 339, 366, 425, 426, 437, 446
hold-out set, 202
Holm's method, 568, 576, 585
hypergeometric distribution, 501
hyperparameter, 187
hyperplane, 367–372
hypothesis test, 76–77, 84, 103, 558–583

IMDb data set, 413, 415, 416, 418, 420, 437, 458, 467
imputation, 515
Income data set, 16–18, 21–23
increment, 60
independent variable, 15
indexable, 186
indicator function, 292
inference, 17, 18
inner product, 379, 380
input layer, 400
input variable, 15
integral, 301
interaction, 70, 89, 95–98, 110–111, 308
intercept, 71, 72
interpolate, 432
interpretability, 229
inversion, 529
irreducible error, 17, 36, 90, 110
iterator, 312

joint distribution, 158

K-means clustering, 11, 521–524
K-nearest neighbors, 135, 164–167
  classifier, 11, 36–37
  regression, 111–115

Kaplan–Meier survival curve, 472–474, 483
kernel, 379–382, 384, 394
  linear, 380
  non-linear, 377–382
  polynomial, 380, 382
  radial, 381–383, 390
kernel density estimator, 159
keyword, 46
Khan data set, 12, 579–581, 583, 590, 593
knot, 290, 294, 296–299

ℓ_1 norm, 244
ℓ_2 norm, 242
lag, 422
Laplace distribution, 251
lasso, 11, 24, 244–251, 265–266, 336, 385, 484
leaf, 333, 526
learning rate, 429
least squares, 5, 21, 71–72, 140, 141, 229
  line, 73
  weighted, 103
level, 92
leverage, 104–106
likelihood function, 141
linear, 2, 69–115
linear combination, 128, 230, 253, 505
linear discriminant analysis, 5, 11, 135, 138, 147–155, 164–167, 377, 382
linear kernel, 380
linear model, 20, 69–115
linear regression, 5, 11, 69–115, 172–173
  multiple, 80–90
  simple, 70–80
link function, 172, 173
linkage, 529–530, 548
  average, 529–530
  centroid, 529–530
  complete, 526, 529–530
  single, 529–530
list, 41
list comprehension, 123
local minimum, 427
local regression, 290
log odds, 145
log-rank test, 474–476, 483
logistic function, 139
logistic regression, 5, 11, 25, 135, 138–144, 164–167, 172–173, 308–309, 377, 384–385
  multinomial, 145, 163
  multiple, 142–144
logit, 140
loss function, 300, 385
low-dimensional, 262
LSTM RNN, 420

main effects, 96
majority vote, 344
Mallow's C_p, 87, 231, 232, 236–238
Mantel–Haenszel test, 475
margin, 370, 385
marginal distribution, 158
Markov chain Monte Carlo, 353
matrix completion, 515
matrix multiplication, 10
maximal margin
  classifier, 367–372
  hyperplane, 370
maximum likelihood, 139–141, 143, 170
mean squared error, 28
mesh, 53
method, 43
minibatch, 429
misclassification error, 35
missing at random, 515
missing data, 56, 515–520
mixed selection, 87
MNIST data set, 402–404, 406, 430, 431, 441, 444, 445, 448
model assessment, 201
model selection, 201
module, 42
multicollinearity, 108, 266
multinomial logistic regression, 145, 163
multiple testing, 557–583
multi-task learning, 403
multivariate Gaussian, 150

multivariate normal, 150

naive Bayes, 135, 158–161, 164–167
namespace, 116
natural spline, 297, 298, 301, 317
NCI60 data set, 4, 5, 12, 546, 548–550
negative binomial, 173
negative predictive value, 155, 156
neural network, 5, 399
node
  internal, 333
  purity, 337–339
  terminal, 333
noise, 21, 252
non-linear, 2, 11, 289–329
  decision boundary, 377–382
  kernel, 377–382
non-parametric, 20, 22–23, 111–115, 193
normal (Gaussian) distribution, 146, 147, 150, 172, 476, 561
notebook, 40
null, 152
  distribution, 561, 578
  hypothesis, 76, 559
  model, 87, 231, 245
null rate, 186
NYSE data set, 12, 422–424, 466, 467

Occam's razor, 426
odds, 140, 145, 195
OJ data set, 12, 365, 398
one-hot encoding, 92, 126, 403
one-standard-error rule, 240
one-versus-all, 384
one-versus-one, 384
one-versus-rest, 384
optimal separating hyperplane, 370
optimism of training error, 30
ordered categorical variable, 315
orthogonal, 257, 506
  basis, 125
out-of-bag, 345
outlier, 103–104
output variable, 15
over-parametrized, 465
overdispersion, 172
overfitting, 21, 23, 25, 30–31, 88, 152, 233, 371

p-value, 77, 82, 560–562, 578–579
  adjusted, 586
package, 42
parameter, 71
parametric, 20–22, 111–115
partial least squares, 254, 260–262, 282
partial likelihood, 480
path algorithm, 249
permutation, 578
permutation approach, 577–582
perpendicular, 257
Poisson distribution, 169, 172
Poisson regression, 135, 167–173
polynomial
  kernel, 380, 382
  regression, 98–99, 289–292, 294–295
pooling, 410
population regression line, 73
Portfolio data set, 12
positive predictive value, 155, 156
posterior
  distribution, 251
  mode, 251
  probability, 147
power, 108, 155, 563
precision, 155
prediction, 17
  interval, 90, 110
predictor, 15
principal components, 505
  analysis, 11, 254–260, 504–515
  loading vector, 505, 506
  missing values, 515–520
  proportion of variance explained, 510–515, 547
  regression, 11, 254–260, 280–282, 504, 515
  score vector, 506
  scree plot, 514–515
prior
  distribution, 251
  probability, 146
probability density function, 477, 478

projection, 230
proportional hazards assumption, 478
pruning, 336
  cost complexity, 336
  weakest link, 336
Publication data set, 12, 482–487
Python objects and functions
  %%capture, 458
  iloc[], 58
  loc[], 57
  AgglomerativeClustering(), 543
  anova(), 313
  anova_lm(), 125, 129, 312, 313
  axhline(), 122, 551
  axline(), 121, 129, 329
  BART(), 362
  biplot, 537
  boot_SE(), 223
  boxplot(), 62, 66
  bs(), 315, 327
  BSpline(), 315
  clone(), 222
  columns.drop(), 122
  compute_linkage(), 544
  confusion_table(), 176
  contour(), 50
  corr(), 129, 174
  cost_complexity_pruning_path(), 357
  CoxPHFitter(), 491
  cross_val_predict(), 270
  cross_validate(), 218, 219, 226
  cumsum(), 539
  cut_tree(), 545
  data.frame(), 227
  Dataset, 440
  decision_function(), 392
  DecisionTreeClassifier(), 354, 355
  DecisionTreeRegressor(), 354
  def, 121
  dendrogram(), 544
  describe(), 62, 66
  dir(), 116
  drop(), 179
  dropna(), 56, 268, 461
  DTC(), *see* DecisionTreeClassifier()
  DTR(), *see* DecisionTreeRegressor()
  dtype, 43
  ElasticNetCV(), 279
  enumerate(), 217
  export_text(), 356
  export_tree(), 365
  fit(), 118, 181, 218
  fit_transform(), 119
  for, 59
  GaussianNB(), 182
  GBR(), *see* GradientBoosting-Regressor()
  get_dummies(), 461
  get_influence(), 121
  get_prediction(), 120, 314
  get_rdataset(), 535
  glm(), 313
  glob(), 437
  GradientBoostingClassifier(), 361
  GradientBoostingRegressor(), 354, 361
  GridSearchCV(), 276
  groupby(), 490
  hist(), 62
  iloc[], 58, 59
  import, 42
  imshow(), 50, 449
  ISLP.bart, 362
  ISLP.cluster, 544
  json, 437
  KaplanMeierFitter(), 502
  keras, 437
  KFold(), 219
  KMeans(), 542, 543
  Kmeans(), 542
  KNeighborsClassifier(), 183
  lambda, 58
  LDA(), *see* LinearDiscriminant-Analysis()
  legend(), 132
  lifelines, 490
  LinearDiscriminantAnalysis(), 174, 179
  LinearGAM(), 317
  LinearRegression(), 280

load_data(), 117
loc[], 58, 59, 177
log_loss(), 355
LogisticGAM(), 323
logrank_test(), 490
lowess(), 324
matplotlib, 48
max(), 66
mean(), 48
median(), 197
min(), 66
MNIST(), 444
ModelSpec(), 116–118, 122, 124, 267
MS(), *see* ModelSpec()
mult_test(), *see* multipletests()
multipletests(), 586
multipletests(), 583, 589
multivariate_logrank_test(), 496
NaturalSpline(), 317, 319
ndim, 42
nn.RNN(), 461
normal(), 132, 286, 555
np, *see* numpy
np.all(), 54, 180
np.allclose(), 190
np.any(), 54
np.arange(), 51
np.argmax(), 122
np.array(), 42
np.concatenate(), 133
np.corrcoef(), 46, 554
np.empty(), 224
np.isnan(), 268
np.ix_(), 53
np.linalg.svd(), 539
np.linspace(), 50
np.logspace(), 318
np.mean(), 47, 176
np.nan, 60
np.nanmean(), 541
np.percentile(), 228
np.power(), 219
np.random.choice(), 553
np.random.default_rng(), 46, 47
np.random.normal(), 45
np.sqrt(), 45
np.squeeze(), 457
np.std(), 47
np.sum(), 43
np.var(), 47
np.where(), 180
ns(), 317
numpy, 42, 555
os.chdir(), 55
outer(), 219
pairwise_distances(), 554
pairwise_tukeyhsd(), 587
pandas, 55
params, 175
partial(), 222, 269
PCA(), 280, 537, 540, 554
pd, *see* pandas
pd.crosstab(), 555
pd.cut(), 315
pd.get_dummies(), 314
pd.plotting.scatter_matrix(), 62
pd.qcut(), 314, 315
pd.read_csv(), 55, 556
pd.Series(), 62
Pipeline(), 275
plot(), 48, 61, 356, 490
plot.scatter(), 120
plot_gam(), 321
plot_svm(), 398
PLSRegression(), 282
poly(), 125, 313, 327
predict(), 175, 178, 181, 216, 218, 323, 358
predict_survival_function(), 493
print(), 40
pvalues, 175
pygam, 307, 317
pytorch_lightning, 435
QDA(), *see* QuadraticDiscriminant-Analysis()
QuadraticDiscriminantAnalysis(), 174, 181
random(), 555
RandomForestRegressor(), 354, 360
read_image(), 436
reindex(), 461
reshape(), 43

return, 198
RF(), *see* RandomForestRegressor()
rng, *see* np.random.default_rng()
rng.choice(), 60
rng.standard_normal(), 60
roc_curve(), 392
RocCurveDisplay.from_estimator(), 387
savefig(), 50
scatter(), 49, 61
scipy.interpolate, 315
score(), 218, 461
seed_everything(), 436
set_index(), 57
set_title(), 49
set_xlabel(), 49
set_xscale(), 198
set_ylabel(), 49
set_yscale(), 198
shape, 43
ShuffleSplit(), 219
sim_time(), 495
SimpleDataModule(), 441
SimpleModule.classification(), 446
SimpleModule.regression(), 442
skl, *see* sklearn.linear_model
skl.ElasticNet(), 273, 277
skl.ElasticNet.path, 274
skl.ElasticNet.path(), 273
sklearn, 118, 181
sklearn.ensemble, 360
sklearn.linear_model, 267
sklearn.model_selection, 267
sklearn_selected(), 269
sklearn_selection_path(), 270
sklearn_sm(), 218
skm, *see* sklearn.model_selection
skm.cross_val_predict(), 271
skm.KFold(), 271
skm.ShuffleSplit(), 272
slice(), 51, 462
sm, *see* statsmodels
sm.GLM(), 174, 192, 226
sm.Logit(), 174
sm.OLS(), 118, 129, 174, 319
StandardScaler(), 185, 438, 537, 555
statsmodels, 116, 173
std(), 186
Stepwise(), 269
str.contains(), 59
subplots(), 48
sum(), 43, 268
summarize(), 118, 129, 223, 226
summary(), 119, 322, 587
super(), 440
SupportVectorClassifier(), 387, 389–391, 393
SupportVectorRegression(), 394
SVC(), *see* SupportVectorClassifier()
svd(), 539
SVR(), *see* SupportVectorRegression()
TensorDataset(), 441
to_numpy(), 437
torch, 435
torchinfo, 436
torchmetrics, 436
torchvision, 436
ToTensor(), 444
train_test_split(), 186, 216
transform(), 118, 119
ttest_1samp(), 584
ttest_ind(), 590
ttest_rel(), 587
tuple, 43
uniform(), 555
value_counts(), 66
var(), 536
variance_inflation_factor(), 116, 124
VIF(), *see* variance_inflation_factor()
where(), 355
zip(), 60, 312

q-values, 589
quadratic, 98
quadratic discriminant analysis, 4, 135, 156–157, 164–167

qualitative, 2, 27, 91, 135, 167, 202
  variable, 91–94
quantitative, 2, 27, 91, 135, 167, 202

radial kernel, 381, 383, 390
random forest, 11, 331, 343, 346–347, 354, 360–361
random seed, 46
re-sampling, 577–582
recall, 155
receiver operating characteristic (ROC), 154, 382–383
recommender systems, 516
rectified linear unit, 401
recurrent neural network, 416–427
recursive binary splitting, 334, 337, 338
reducible error, 17, 90
regression, 2, 11, 27
  local, 289, 290, 304–305
  piecewise polynomial, 294–295
  polynomial, 289–292, 299
  spline, 289, 294
  tree, 331–337, 358–360
regularization, 230, 240, 406, 484–486
ReLU, 401
resampling, 201–214
residual, 71, 81
  plot, 100
  standard error, 75, 77–78, 88–89, 109
  studentized, 104
  sum of squares, 71, 79, 81
residuals, 263, 348
response, 15
ridge regression, 11, 240–244, 385, 484
risk set, 473
robust, 374, 376, 535
ROC curve, 154, 382–383, 486–487
R^2, 77–80, 88, 109, 238
rug plot, 314

scale equivariant, 242
Scheffé's method, 572
scree plot, 512, 514–515
  elbow, 514
semi-supervised learning, 27
sensitivity, 153, 155, 156
separating hyperplane, 367–372
Seq2Seq, 425
sequence, 41
shrinkage, 230, 240, 484–486
  penalty, 240
sigmoid, 401
signal, 252
signature, 45
singular value decomposition, 539
slack variable, 375
slice, 51
slope, 71, 72
`Smarket` data set, 2, 3, 12, 173, 184, 196
smoother, 308
smoothing spline, 290, 300–303
soft margin classifier, 372–374
soft-thresholding, 250
softmax, 145, 405
sparse, 244, 252
sparse matrix format, 414
sparsity, 244
specificity, 153, 155, 156
spline, 289, 294–303
  cubic, 296
  linear, 296
  natural, 297, 301
  regression, 289, 294–299
  smoothing, 30, 290, 300–303
  thin-plate, 22
standard error, 75, 101
standardize, 185
statistical model, 1
step function, 111, 289, 292–293
stepwise model selection, 11, 231, 233
stochastic gradient descent, 429
string, 41
string interpolation, 490
stump, 349
subset selection, 230–240
subtree, 336
supervised learning, 25–27, 261
support vector, 371, 376, 385
  classifier, 367, 372–377
  machine, 5, 11, 24, 377–386

regression, 386
survival
analysis, 469–502
curve, 472, 483
function, 472
time, 470
synergy, 70, 89, 95–98, 110–111
systematic, 16

t-distribution, 77, 165
t-statistic, 76
t-test
one-sample, 583, 584, 588
paired, 587
two-sample, 559, 570, 571, 577–581, 584, 590
test
error, 35, 37, 176
MSE, 28–32
observations, 28
set, 30
statistic, 559
theoretical null distribution, 577
time series, 101
total sum of squares, 79
tracking, 102
train, 21
training
data, 20
error, 35, 37, 176
MSE, 28–31
transformer, 311
tree, 331–342
tree-based method, 331
true negative, 155
true positive, 155
true positive rate, 155, 156, 382
truncated power basis, 296
Tukey's method, 571, 585, 587
tuning parameter, 187, 240, 484
two-sample t-test, 474
Type I error, 155, 562–565
Type I error rate, 563
Type II error, 155, 563, 568, 584

unsupervised learning, 25–27, 255, 260, 503–552
`USArrests` data set, 12, 507, 508, 510, 512, 513, 515, 516, 518, 519

validation set, 202
approach, 202–204
variable, 15
dependent, 15
dummy, 91–94, 97–98
importance, 346, 360
independent, 15
indicator, 35
input, 15
output, 15
qualitative, 91–94, 97–98
selection, 86, 230, 244
variance, 18, 31–34, 159
inflation factor, 108–110, 123
varying coefficient model, 305

`Wage` data set, 1, 2, 8, 9, 12, 290, 291, 293, 295, 297–300, 302–306, 309, 315, 327
weak learner, 343
weakest link pruning, 336
`Weekly` data set, 12, 196, 226
weight freezing, 412, 419
weight sharing, 418
weighted least squares, 103, 304
weights, 404
with replacement, 214
within class covariance, 150
wrapper, 217