Index

ACM–SIGKDD Innovation Award, 762
activation functions, 837–842
actual and predicted profit values, 325
adjusted R-squared (adjusted R^2) method, 298–299
agglomerative clustering
  AGNES, 648
  average linking, 653
  bottom-up approach, 648
  clustering of data points, 654
  clustering of students record, 663–675
  complete linkage, 652
  distance matrix, 656–658
  and divisive clustering, 648–649
  graphical representation, 653–654
  HCA diagram, 650–651
  over supermarket dataset, 659–663
  process, 649–650
  single linkage, 651–652
  weaknesses of, 675
AGglomerative NESting (AGNES), 648
aggregation, 138
Alexa Purchase Problem, 554
all possible models, 297
Anaconda Navigator
  Jupyter, 61–62
  Spyder, 60
ANN model
  compilation, 875–876
  initialization, 872
  input and first hidden layer, 872–873
  libraries, 871–872
  output layer, 874–875
  second and third hidden layer, 873–874
  training set, 876–878
applications of ML
  bioinformatics, 7
  chatbots, 4
  dynamic pricing, 3–4
  handwritten text recognition, 5
  healthcare and pharmaceuticals, 6
  image recognition, 2–3
  machine translation software, 5
  malware detection, 2–3
  online fraud detection, 6
  online video streaming platforms, 4
  product recommendations, 6
  search engines, 7
  self-driving cars and autonomous drones, 7
  smart home assistants, 2–3
  social media services, 5
  spam emails, 2–3
  speech recognition, 2–3
  traffic prediction with Google Map, 4–5
  video surveillance, 5–6
Apriori algorithm
  candidate 1-itemsets C1 of database, 763
  confidence of all possible rules, 768

frequency of candidate item pairs, 766
frequent 1-itemsets, 764–765
frequent itemsets identification, 769–770
frequent 1-itemsets L1, 764–765
inventors, 760–762
limitations of, 797
phases, 762
property, 781–791
sample transaction database, 775
TID, 795–800
transactional database, 760, 763
architecture of basic CNN model, 905–914
architecture of RNN
compilation, 1026
3D structure, 1033–1036
first layer of LSTM, 1024–1026
Keras libraries and packages, 1022–1024
output layer, 1026
test dataset, 1029–1033
training set, 1027–1029
artificial intelligence (AI), 46–47
artificial neural network (ANN). *See also* role of weights
architecture of, 972
deep learning neural network, 829
durability and access time of some storage techniques, 823
flow of information, 829
history of storage cost, 822
layer structure, 828
multiple input values, 831–833
processing capacity of computers, 823
squashing, 972
technology, 822
temporal lobe and, 971
working of, 829–831
artificial neuron, 827
association rule mining, 24, 26. *See also* Naive algorithm
advantages of, 743
antecedent and consequent, 744
confidence, 747–749
defined, 743
implementation, 807–819
lift, 748–752
need for, 739–740
rule-based technique, 739
sale of beer and diapers, 740–742
sales database, 744–745
support, 745–746
astronomical data analysis, 26
autonomous drones, 7
average linking, 653
average pooling technique, 919–920

backpropagation through time (BTT), 985
backward elimination method, 293–295
backward propagation process, 845, 860
bi-directional elimination method, 296–297
bioinformatics, 7
biological neuron, 827
Brute-force approach, 852–853

cancer dataset, 866
candidate 1-itemsets C1 of database, 763
casting, 68–69
categorical attributes, 275, 318
categorical data
city and purchased variables, 168
dummy variables, 150
label-encoded data, 149
labeling, 148
processed data, 150–151
separate attributes, 149
categorical variables, 168–169
nominal, 14
ordinal, 15
cell state, 999–1000
chatbots, 4
Chebyshev distance, 620–621
Chervonenkis, Alexey, 506
Chi-square distribution table, 284–292
city variable, 168
classification. *See also* decision tree classifier
binary and multi-class, 361
building classifier, 365–366
decision tree classifier, 364
defined, 360

examples of regression, 360–361
input and output attributes, 362–363
lazy and eager learner classifiers, 367
multi-label, 361
posteriori, 362
priori, 362
training and testing sets, 365–366
training dataset, 367
two-step process, 363
classification algorithms
decision tree classifier, 544–554
pre-processing data template, 534–544
stepwise approach, 534–535
classification of ML algorithms
supervised learning, 15–23
unsupervised learning, 23–26
classifiers of SVMs
history of, 506
linear SVM classifier, 506
non-linear SVM classifier, 507
closed and maximal itemsets, 791–795
clustering. *See also* distance metrics; partitioning clustering
applications of, 613–614
characteristics of, 612
defined, 611–612
methods/algorithms, 621–622
model, 612
properties, 614–615
values of attributes, 612
clustering/outlier analysis, 136
ColumnTransformer constructor, 319
comment in Python, 65
compilation of CNN, 946–947
complete linkage, 652, 655, 658
complex curves, 309–310
concept hierarchy generation, 142
confusion matrix, 480–481, 484–485, 547–548, 557–559, 880–883
continuous variables, 15
contour plots
after taking tiny step, 210
analysis of cost function, 207–208
color ring, 205
cost function and parameters, 204
defined, 204
gradient descent, 214
hypothesis plot, 205
local minima for same plot, 212
minimum value of, 206
with one parameter, 213
slope of curve, 219–224
slope of line, 214–219
converting image to numbers, 893–896
convolutional neural network (CNN), 889
computer perceives image, 904
convolution layer, 906–914
2D array, 904
3D array, 905
occipital lobe and, 971
pixel matrix representation, 905
typical image classification task, 903
convolution layer, 906–914
CNN model, 941
input image processing, 941
and max pooling layer, 954–957
process, 942
cost function
contour plots, 204–208
defined, 198
3D graph of, 203
with one parameter, 199–203
covariance matrix, 144–145
customers of mall clustering
customer's Annual Income and Spending Score, 701
dataset, 700–701
dendrogram, 712–714
elbow method, 704–707
k-means, 707–709
libraries and dataset, 702–704
SuperDataScience dataset, 701
visualization, 709–712

data cleaning
data inconsistency, 137
description, 130
missing values, 130–133

noisy data, 133–137
dataframe dataset, 245
data integration
description, 137
metadata, 138
redundancy of data, 138
data mining process, 45
data pre-processing. *See also* data cleaning; data integration; data reduction; data transformation
garbage in and garbage out, 128–129
house price prediction, 125–126
precision agriculture applications, 127–128
stages of, 129
data reduction
feature extraction, 143
feature selection, 143
data science, 45
dataset and pre-processing
categorical data, 869–870
dependent and independent variables, 867–869
feature scaling, 871
libraries, 866–867
training and testing, 870
data transformation
aggregation, 138
concept hierarchy generation, 142
discretization, 142
feature scaling, 138–140
feature selection, 140–142
data types in Python, 67
decision tree classifier. *See also* information gain; random forest classification model; SVM linear model
accuracy, 548–549
advantages and disadvantages, 445
building, 370
classification model and predictions, 544
comparative analysis, 592–594
confusion matrix, 547–548
correlation of, 368
if–then statements, 369
information theory, 370–373
performance metrics and visualization, 546–547
precision, 549
prediction of, 369
recall, 549–550
root nodes, 373–374
structure of, 367–368
test set prediction, 545–546
training set, 544–545
visualization of test results, 553–554
deep learning neural network, 45, 829
definitions of ML, 10–11
dendrogram, 648
description, 712
mall customers, 712–713
optimal number of clusters, 713–714
density-based clustering
arbitrary shapes and noise, 681–682
border data point, 684
core data point, 683
core point, 683–684
DBSCAN, 682
density-connectivity, 687–688
directly density-reachability, 685
indirectly density-reachability, 685–686
k-means, 681
MinPts, 682–683
neighborhood (ε), 682
noise data point/outlier, 684
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
algorithm, 688–689
defined, 682
disadvantages, 690–692
strengths of, 690
density-connectivity concept, 687–688
directly density-reachability, 685–686
discrete variables, 15
distance matrix, 675–679
distance metrics
Chebyshev distance, 620–621
Euclidean distance, 615–617
Manhattan distance, 618–619
Minkowski distance metric, 619–620
proximity measure, 615
DIvisive ANAlysis (DIANA), 649

divisive clustering
    dendrogram of, 680
    DIANA, 649
    distance matrix, 676–679
    HCA diagram, 675
    record of students' performance, 675–676
    top-down approach, 649, 675
document analysis, 613
Dog–Cat classifier
    CNN model, 940–941
    convolution layer, 941–943
    description, 938
    directory structure, downloaded, 938
    flattening, 944–945
    full connection, 945–946
    keras library and packages, 940
    phases of CNN model, 939
    pooling, 943–944
    training and testing set, 939
double-point crossover, 1068–1071
dummy variables, 150
    Boolean indicator, 271–272
    explanation of, 320
    multiple linear regression, 272, 274
    and nominal attribute, 273
    nominal attribute replacement, 272
    populated table, 273
dynamic pricing, 3–4

eigenvectors and eigenvalues, 145–147
elbow method, 646–647
    KMeans() constructor, 705
    number of clusters, 706
    optimal number of clusters, 704–707
elitism
    defined, 1076
    and elitism ratio, 1077–1079
    working of operators, GA, 1075–1076
elitism ratio, 1077–1079
epoch, 845
error of prediction, 196
estate price prediction problem, 846, 848
Euclidean distance, 178, 615–617

fake news identification, 614
feature extraction, 143
feature scaling
    attributes/columns, 178
    Euclidean distance, 178
    normalization, 179–181
feature selection, 143
    customer dataset, 141
    defined, 140
    dropping columns, 141
    filters, 141–142
    wrappers, 141
features of Python, 54–55
feed-forward networks
    classification at time, 975
    image classification, 975
    prediction of, 976
50-degree polynomial regression plot, 342
filters, 141–142
fine-tuning weights, 924
fit() method, 249–250
fitness function, 1053–1054
flattening operation, 920–922, 944–945
forget gate, 997–998
for loop, 107–110
forward propagation process, 843, 859–860
forward selection method, 295–296
fraud detection, 613
frontal lobe and RNN, 971
full connection, 922–923, 945–946

gamma parameter, 526–527
Gaussian Kernel Function, 516–519
general AIs, 47
generation of synthetic data, 897–898
genetic algorithm (GA). *See also* elitism; role of recombination
    advantages, 1079
    applications, 1080
    chromosome, 1049
    constraints, 1050
    Darwin's theory, 1042–1043
    description, 1041
    disadvantages, 1080
    fitness function, 1053–1054
    gene, 1049

gradient descent algorithm, 1042
mutation process, 1046–1049
natural selection, 1041, 1043–1044
objectives, 1050–1051
population, 1049
recombination/crossover process, 1044–1046
representation, 1052–1053
search space, 1051–1052
variables, 1050
Gini index, 408–444
Godfathers of Deep Learning, 829
Google live albums, 891
Google Neural Machine Translation (GNMT), 5
Google Photos app, 891
gradient descent algorithm, 853–856, 1042
basic idea, 208–209
3D plot, cost function, 209
initialization of parameters, 209–210
values of parameters, 229–235
graphical user interface (GUI), 60

handwritten digit classifier, 958–965
handwritten text recognition, 5
hidden layer neurons, 851, 945
hierarchical clustering algorithms (HCA)
agglomerative clustering. *See also* agglomerative clustering
defined, 648
dendrogram, 648
density-based clustering. *See also* density-based clustering
divisive clustering. *See also* divisive clustering
history of ML
contemporary phase (post 2000), 10–11
foundational phase (before 1940), 8
phases of development, 7–8
transitional phase (1940–2000), 9
history of Python, 55–56
Horizontal storage, 758–759
human brain. *See also* artificial neural network (ANN); convolutional neural network (CNN); recurrent neural network (RNN); three-step learning process
biological and artificial neuron, 827
identification of images, 900–903
input attributes, 828
network of neurons, 825–827
processing signals, 824
structure of, 970
training of neural network, 842–852
hyperbolic tangent function, 840–842
hyperparameter (σ), 519–520
hyperplane, 499–503
Euclidean spaces, 514
maximized margin, 504
positive and negative hyperplane, 505–506
support vectors, 504
hypothesis/function, 194

identifiers in Python, 63
if statement, 102–105
iloc (Pandas selection and indexing), 245–246
image augmentation, 947–951
image recognition, 2–3, 889–892
implementation of RNN
data pre-processing
dataset reshaping, 1020–1022
data structure, 1015
DelhiDieselPrice dataset, 1008–1009
downloaded dataset, 1009
feature scaling, 1013–1014
histogram, 1010–1013
independent variables, 1016–1017
training and testing data, 1010
indentation in Python, 64
independent rank and dependent salary variable, 331
indirectly density-reachability, 685–686
information gain
"age" attribute, 376–377, 383–384, 388–389
check trend of profit as up or down, 375–376
"competition" attribute, 377–378, 381–382
"credit rating" attribute, 386–387, 390–391
defined, 374
Gini index, 408–444
"income" attribute, 384–385, 389–390, 392–408

- “mid” attribute, 380–381
- “new” attribute, 380
- “old” attribute, 379
- “student” attribute, 385–386
- “type” attribute, 378–379
- “yes” attribute, 387–388

information theory, 370–373
input and output attributes, 362–363
input dimension of images, 953
input gate, 998–999
installation of Python, 56–57
institutional position dataset, 301
insurance policy-holders, 613
integrated development environment (IDE), 60
intelligent personal assistants, 2–3
interactive mode programming, 58
interactive Python, 62
IPython notebook, 62
Iris dataset, 729–735
Iterative Dichotomiser 3 (ID3), 370

joining tuples, 87
Jupyter IDE, 62

keras library and packages, 940
kernel trick
- defined, 515
- 3D perspective, 521–524
- multiple kernels to dataset, 525–526
- RBF/Gaussian kernel function, 516–519

keywords in Python, 63–64
k-means clustering algorithm
- customers of mall, 700–719
- database after first iteration, 626–627
- database after initialization, 625
- database after second iteration, 628–630
- data plot, 630
- defined, 699–700
- Euclidean distance, 637–640
- example database, 624, 625
- final allocation, 644–645
- first iteration–allocation, 642
- flowchart for, 624
- implementation, 630–633
- implementation and issues, 647–648
- input records, 622–623
- intra cluster distance, records in cluster, 645
- Mall–Customer dataset, 633–634
- optimal number of clusters, 646–647
- record of students’ performance, 641
- second iteration–allocation, 643–644
- seed records, 634–635, 641
- Spyder IDE, 700
- updated centroids, 643
- visualization, 716–719
- working of, 623

k-Nearest Neighbor Algorithm, 467–470
k-NN model, 567–574
knowledge-based method, 293

Laplace estimator, 466–467
lazy and eager learner classifiers, 367
L1-distance/taxicab norm, 618–619
learning rate α, 224–229
level plots, 204
limitations of RNN
- long-term dependencies, 984–985
- short-term dependencies, 983–984

linear regression with one variable, 195
linear SVM classifier, 506
logistic regression classifier model, 470–478, 574–581
long short-term memory (LSTM)
- caption generation from image, 1001–1002
- cell, 995–996
- cell state, 999–1000
- classification of DNA sequence, 1002–1003
- defined, 989
- forget gate, 997–998
- gated units, 990
- hidden state flow between cells, 991–992
- input gate, 998–999
- machine-readable vectors, 991
- machine translation, 1002
- named entity recognition, 1002
- online review of laptop, 990
- output gate, 1000–1001
- sentiment classification, 1003

sigmoid activation function, 996–997
speech recognition, 1002
tanh activation function, 993–995
text generation/language modeling, 1001
time series analysis, 1003
video activity recognition, 1002

machine-readable vectors, 991
machine translation software, 5
Mall–Customer dataset, 633–634
Manhattan distance, 618–619
market basket analysis, 807
market segmentation, 25
Markov's decision process, 34–35
mathematical modeling and RNN, 980–983
Matplotlib, 2D plotting library, 158
"Maximal Margin Classifier" model, 506
maximized margin hyperplane, 504–506
max pooling technique, 916–919
mean absolute error (MAE), 257–258, 325
mean squared error (MSE), 198, 258–259, 325
mini-batch method, 858–859
Minkowski distance metric, 619–620
Min–Max scaling, 139, 179
missing values, 166–168
mean/median/mode, 131–132
observation, 133
prediction of, 132–133
rows/columns deletion, 130–131
Mitchell, Tom M., 11
MNIST dataset, 958–965
Modified National Institute of Standards and Technology (MNIST) dataset, 892–893
multiple input values, 831–833
multiple linear regression
dataset of companies, 268–269
dependent and independent variables, 269–270
instance of 50_AdAgency, 314–316
simple and, 267, 268
multiple regression model, 275–278
dependent variable, 276
garbage in–garbage out, 277
selection of attributes, 277
too complex model, 277
mutation process
abnormalities, 1048
defined, 1046
effect on human beings, 1047
mathematical modeling, 1072–1074
natural selection and recombination, 1047
probability of mutation, 1074–1075

naive algorithm
limitations of, 755–756
non-zero frequencies transaction dataset, 756–757
sale record of grocery store, 753
working of, 753–755
naive Bayes classifier model
Alexa Purchase Problem, 561
classification model and predictions, 562–563
performance metrics and visualization, 563–565
visualization of test set, 566–567
visualization of training, 565–566
naive Bayes method
for classification, 451–466
defined, 448
Laplace estimator, 466–467
manufactured screws, 449–451
ML and AI classification, 448
named entity recognition, 1002
National Institute of Industrial Engineering (NITIE), 761
natural selection, 1041, 1043–1044
Roulette wheel working, 1054–1061
noisy data
clustering/outlier analysis, 136
data cleaning tasks, 133
regression, 136–137
smoothing/binning methods, 133–135
nominal attributes, 318
handling, 270–271
separate, 271
with three values, 273
nominal variables, 14
non-linear SVM classifier, 507

Cambridge University Press & Assessment
978-1-009-17024-6 — Machine Learning with Python
Parteek Bhatia
Index
More Information

non-linear data, 508
non-linearly separable data, 507–508
one-dimensional (1D) dataset, 509–513
TRIANGLES and PENTAGON classes, 508–509
two-dimensional (2D) dataset, 513–514
non-linear SVM linear model, 587–592
normalization, 139–140
input variables, 179
limitation of, 180–181
in Python, 179–180
Not A Number (NaN), 989
null hypothesis, 292–293
numerical variables
continuous, 15
discrete, 15
NumPy/Numeric Python, 156–157

occipital lobe and CNN, 971
OneHotEncoder, 319
one-hot encoding, 170–175
online video streaming platforms, 4
optimal hyperplane, 504
optimal number of epochs, 925–930
optimal weights, 845
ordinal variables, 15
output gate, 1000–1001
output layer, 945–946
overfitting model, 926–928

Pandas library, 157
parietal lobe and ANN, RNN, 971–972
partitioning clustering
defined, 622
k-means (*see* k-means clustering)
plot of salary *versus* rank, 302
polynomial linear regression
behavior of COVID confirmed cases, 299–300
chemistry lab experiment, 348–355
complex curves, 309–310
cubic relationship, 309
dataset import, 329
defined, 327
independent variables, 303–305
institutional position dataset, 328
linear, 301
linear relationship, 309
mathematical representation, 300
multiple input attributes, 308
quadratic relationship, 309
rank–salary problem, 302–303
vs. simple linear regression, 344–347
variable explorer, 329
pooling layer, 915–916, 943–944
positive and negative hyperplane, 505–506
Posteriori classification, 362
predict() method, 250
principal component analysis (PCA)
algorithm, 148
concept of, 147
covariance matrix, 144–145
description, 143
eigenvectors and eigenvalues, 145–147
standardization process, 144
priori classification, 362
Purchase Alexa Dataset, 533–534, 536
purchased variable, 168
P-value (probability value)
actual observed results, 281
Chi-square distribution table, 284
Chi-square value, 282–283
degree of freedom, 281–282
expected results, 280–281
experiment, 280
null hypothesis, 278–279
significance level, 279
value of significance level, 283
Python dictionary, 88–95
Python functions, 110–114
Python list
description, 73–74
features of, 74
operations on, 74–81
Python numbers, 68–69
Python set, 96–101
Python strings, 69–73

Q-learning
Q-matrix, 37–44
reinforcement learning algorithm, 35
representation house, 35–36
reward matrix, 36–37
Q-matrix, 37–44
quality of classifier
confusion matrix, 480–481
defined, 478
false negative, 480
false positive, 480
F-measure, 483
precision, 481–482
recall, 482–483
true negative, 479
true positive, 479
Quinlan, John Ross, 370

radial basis function (RBF), 516–519
random forest classification model
Alexa Purchase Problem, 554
confusion matrix, 557–559
fitting training set, 555–556
performance metrics and visualization, 557
principle of ensemble learning, 446
test set predictions, 556–557
visualization of test set, 560–561
visualization of training, 560
working of, 446–447
random selection, 1064
ranking of computer programming languages, 53–54
rank selection, 1063–1064
recombination/crossover process, 1044–1046
rectified linear unit (ReLU) activation function, 915
rectifier linear unit function, 839–840
recurrent neural network (RNN). *See also* limitations of RNN
backpropagation, 985
case study of coaching center, 977–980
exploding gradient, 988–989
frontal lobe and, 971
mathematical modeling, 980–983
single arrow between layers, 973
temporal loop, 973–974
vanishing gradient, 986–988
vertical representation of layers, 973
working inside cell, 992–993
working principle of, 976–977
regression metrics, 136–137
evaluation, 257
MAE, 257–258
MSE, 258–259
RMSE, 259–260
regularization parameter, 527–528
reinforcement learning. *See also* Q-learning
defined, 27
exploitation, 33
exploration, 32
Markov's decision process, 34–35
reward maximization and discount, 31–32
SAR triples, 33
vs. supervised and unsupervised, 28
working principle of, 28–31
right-fitted model, 928
role of recombination
defined, 1065
double-point crossover, 1068–1071
single-point crossover, 1065–1067
role of weights, 833–835
root mean squared error (RMSE), 259–260, 325
root nodes, 373–374
R-squared (R^2), 260–264

Samuel, Arthur, 10–11
scatter() method, 252, 333
Scikit–Learn Library, 158
script mode programming, 59–60
seed records, 634–635, 641
self-driving cars, 7
sentiment classification, 1003
Shannon, Claude, 371
sigmoid activation function, 838–389, 996–997
simple image classifier, 896–897
simple linear regression (SLR), 986

datasets, 188
defined, 187
equation of straight line, 190
fitting straight line to model, 192
five years of research experience, 192
mathematical equation, 193
notations, 193
parameters, 195–197
ResearchExperience and Stipend, 241–243
research experience *vs.* stipend plot, 189
stipend value, 190–191
training set, 188
simple transaction storage, 758
single linkage, 651–652, 659, 663
single-point crossover, 1065–1067
sliding window approach, 898–899
smart home assistants, 2–3
smoothing/binning methods
bin boundary, 134–136
bin mean, 134–135
bin median, 134–135
datasets, 134
discretizing, 133
social media services, 5
social network analysis, 25
spam emails, 613
speech recognition, 2–3, 1002
Spyder software, 60
dataset data frame, 161–162
dataset import, 161
File Explorer, 159
File explorer, 243
IDE, 61
libraries, 244
row and column selection, 162–164
save and run file options, 160
variable explorer, 244
standardization, 140
process, 144
in Python, 181–182
stationary kernel, 516–519
stochastic gradient descent algorithm, 856–858
stochastic universal sampling (SUS), 1062
sum of square errors, 196
supervised learning
considered label data, 16
data plot, 20
datasets, 19
defined, 15–16
features for classification, 21
prediction of cost of plot, 17–18
predictions, 20
price prediction dataset, 16–17
regression and classification, 22
vs. unsupervised, 26–27
support vector machines (SVMs). *See also* classifiers of SVMs; hyperplane; kernel trick; tuning parameters
dataset distribution, 496
hyperplane separation, 497–499
ML models, 495
terminologies, 503
SVM linear model, 581–587

tanh activation function, 993–995
temperature and vapor pressure readings, 306
temporal lobe and ANNs, 971
text classification, 594–604
text generation/language modeling, 1001
three-class classifier, 484–485
three-step learning process, 835–842
threshold function, 837–838
time series analysis, 1003
tournament selection, 1063
traditional programming *vs.* ML, 13–15
training and testing datasets, 175–177, 246–249
training set, 188
train_test_split() method, 247
transaction database storage
horizontal, 758–759
simple, 758
vertical representation, 759–760
tuning parameters
defined, 526
gamma parameter, 526–527
regularization parameter, 527–528
tuples
data type, 81–82

features of, 82
and lists, 87–88
operations over, 83–87

underfitting model, 926
univariate regression, 187
unsupervised learning
anomaly detection, 26
applications of, 24
association rule mining, 24, 26
astronomical data analysis, 26
computing clusters, 25
defined, 23
Google News, 24–25
market segmentation, 25
social network analysis, 25
vs. supervised, 26–27
working of, 23
updated distance matrix, 661–662, 668–674
U.S. arrests dataset
clusters visualization, 724–726, 728–729
dendrogram, 726–727
HCM, 727
instance of, 719–720
k-means, 722–724
libraries and dataset, 720–721
optimal number of clusters, 721–722

Vapnik, Vladimir, 506
variable explorer tab, 250, 317
variables in Python
declaring and assignment of values, 66–67
defined, 65
video activity recognition, 1002
video surveillance, 5–6
virtual personal assistants, 2–3

weak AI systems, 47
while loop, 105–107
wrappers, 141

X and Y matrices, 165–166

Yann LeCun, 892

zig-zag curve (overfitted model), 927