Index

@ syntax, 76
__class__ attribute, 114
__init__() function, 108, 121
__init__.py file, 42
__name__ attribute, 62, 179

3D structure. See structures

active substitutions, 275
alignments
amino acid properties, 257
BLAST, 225
ClustalW, 241
conservation, 255
gap penalties, 219
guide tree, 269
mapped genomic, 357
multiple, 232, 241, 269
pairwise, 221
profile, 235–236
alternative hypothesis, 456
amino acids, 182
AND operation, 49
arithmetic operations, 25
arrays
hierarchical clustering, 338
normalisation, 328
NumPy, See NumPy arrays
quantile normalisation, 333
reading image files, 321
text files, 319
value clipping, 373
writing text files, 325
Z-score normalisation, 330
atan2() function, 139
attributes
checking, fetching, 113
frozen, 136
getters, setters, 132
listing, 113
privacy, 134
bar charts, 142
Bayes’ theorem, 429
best-fit line, 485
binary operations, 613
binomial distribution, 431
binomial test, 464
BioPython, 205
multiple alignments, 242
reading sequence files, 205
structures, 312

writing sequence files, 206
BLAST, 225
BLOSUM matrix, 216
bool() function, 49
Boolean data type, 24, 47
Bowtie program, 349
break statement, 53

C language
calling functions, 603
Cython, 598
interface, 587
modules, 591
Python library, 594
Python values, 596
calling external programs, 225, 242
cardinality, 122
catching exceptions, 58
cell counting, 378
central limit theorem, 469
classification
linear discriminant analysis, 507
neural network, 526
support vector machine, 536
ClustalW, 241
clustering, 489
False value, 47
Fast Fourier transform, 388
FASTA sequence file format, 86, 93, 205
FASTQ files, 351
quality scores, 352
reading, 356
feature vector, 487
feed-forward neural network, 523
files
accessing, 81
character-delimited fields, 95
compression formats, 348
copying, 641
CSV, TSV formats, 96
deleting, moving, 98
directories, paths, 98
descriptor, characters, 92
dictionary, existence, 98
FASTA format, 86, 93, 205
FASTQ format, 351, 356
file object, 81, 632
GFF format, 358
GZIP compressed, 348
handles, 81
line iteration, 83
opening, 81
PDB format, 129, 287
pickling, serialisation, 99
reading, 82
SAM format, 357
whitespace delimited, 84
writing, 92
XML format, 90
finally statement, 60
Fisher's linear discriminant, 507
floating point data type, 24
for...else, 53
for loops, 51
format() function, 654
formatting text. See strings: formatting forward-backward algorithm, 447
Fourier transform, 385
frozen sets, 34
FTP downloads, 345
functions
*, ** arguments, 70
arbitrary, anonymous arguments, 70
arguments, 66-67
calling, 64
decorators, 76
defaulting arguments, 67
defining, 64
in classes, 103
keyword dictionaries, 175
lambda, 74
named arguments dictionary, 72
nesting, 74
returning output, 66
gap penalty, 219
Gaussian distribution, 467
tailed test, 469
generator, 171
 genetic code, 186, 190
gene index, 349
genomic alignment, 351
genomic feature data, 358
geometric distribution, 436
getattr() function, 113
GFF annotation files, 359
global statement, 73
graphics
images, 364
user interfaces, 568
widget objects, 570, 578
graphs, 141
greyscale colour model, 362
G-test, 479
GUIs, 566
gzip module, 348
hasattr() function, 113
hash maps. See dictionaries
hierarchical clustering, 338
HMMs, 442
HSV colour model, 362
HTSeq module, 355
if statement, 46, 51
ternary operator, 167
images
blurring, 376
colour components, 363
colour models, 362
contrast adjustment, 370
convolution, 374
counting cells, 378
cropping, 365
detecting, 376
definitions, 373
gamma correction, 371
Gaussian filter, 376
intensity histogram, 373
NumPy arrays, 367
object representation, 364
PIL module, 364
pixmaps, 361
reading from file, 364, 368
resizing, 365
rotating, 365
scaling, 365
sharpening, 370, 375
Sobel filter, 377
thumbnails, 366
value clipping, 373
value normalisation, 371
writing to file, 365

© in this web service Cambridge University Press
www.cambridge.org
<table>
<thead>
<tr>
<th>Keyword</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>import statement</td>
<td>41</td>
</tr>
<tr>
<td>importing modules</td>
<td>41</td>
</tr>
<tr>
<td>indentation</td>
<td>45</td>
</tr>
<tr>
<td>indexing</td>
<td>35</td>
</tr>
<tr>
<td>NumPy arrays</td>
<td>151</td>
</tr>
<tr>
<td>strings</td>
<td>27</td>
</tr>
<tr>
<td>tuples</td>
<td>35</td>
</tr>
<tr>
<td>installation</td>
<td>18</td>
</tr>
<tr>
<td>integer data type</td>
<td>24</td>
</tr>
<tr>
<td>in keyword</td>
<td>176</td>
</tr>
<tr>
<td>is versus ==</td>
<td>48, 176</td>
</tr>
<tr>
<td>isinstance() function</td>
<td>115</td>
</tr>
<tr>
<td>joining</td>
<td>98</td>
</tr>
<tr>
<td>file paths</td>
<td>30, 167</td>
</tr>
<tr>
<td>strings</td>
<td>167</td>
</tr>
<tr>
<td>keywords</td>
<td>607</td>
</tr>
<tr>
<td>k-means clustering</td>
<td>496</td>
</tr>
<tr>
<td>k-nearest neighbours</td>
<td>515–516</td>
</tr>
<tr>
<td>Kohonen map</td>
<td>518</td>
</tr>
<tr>
<td>Kullback-Leibler divergence</td>
<td>198–199, 479</td>
</tr>
<tr>
<td>lambda statement</td>
<td>74</td>
</tr>
<tr>
<td>leading whitespace</td>
<td>45</td>
</tr>
<tr>
<td>length of lists, tuples</td>
<td>36</td>
</tr>
<tr>
<td>line graphs</td>
<td>141</td>
</tr>
<tr>
<td>linear regression</td>
<td>484</td>
</tr>
<tr>
<td>lists</td>
<td>32</td>
</tr>
<tr>
<td>adding items</td>
<td>174</td>
</tr>
<tr>
<td>changing items</td>
<td>36</td>
</tr>
<tr>
<td>comprehension</td>
<td>55</td>
</tr>
<tr>
<td>copying</td>
<td>168</td>
</tr>
<tr>
<td>filtering</td>
<td>56</td>
</tr>
<tr>
<td>manipulation</td>
<td>34</td>
</tr>
<tr>
<td>matrices</td>
<td>148</td>
</tr>
<tr>
<td>membership</td>
<td>35</td>
</tr>
<tr>
<td>operations</td>
<td>621, 625</td>
</tr>
<tr>
<td>reversing</td>
<td>37, 168</td>
</tr>
<tr>
<td>sorting</td>
<td>37, 172</td>
</tr>
<tr>
<td>literal</td>
<td>23</td>
</tr>
<tr>
<td>log() function</td>
<td>138</td>
</tr>
<tr>
<td>logical connectives</td>
<td>49</td>
</tr>
<tr>
<td>loops</td>
<td>53</td>
</tr>
<tr>
<td>break statement</td>
<td>53</td>
</tr>
<tr>
<td>continue statement</td>
<td>52</td>
</tr>
<tr>
<td>dictionary keys</td>
<td>57</td>
</tr>
<tr>
<td>for</td>
<td>51</td>
</tr>
<tr>
<td>indices</td>
<td>54</td>
</tr>
<tr>
<td>list comprehension</td>
<td>55</td>
</tr>
<tr>
<td>skipping, stopping</td>
<td>52</td>
</tr>
<tr>
<td>while</td>
<td>52</td>
</tr>
<tr>
<td>macromolecular structure</td>
<td>286</td>
</tr>
<tr>
<td>philosophy</td>
<td>278</td>
</tr>
<tr>
<td>Markov chains</td>
<td>438</td>
</tr>
<tr>
<td>bacterial growth example</td>
<td>439</td>
</tr>
<tr>
<td>discrete-time homogeneous</td>
<td>438</td>
</tr>
<tr>
<td>forward-backward algorithm</td>
<td>447</td>
</tr>
<tr>
<td>HMMs</td>
<td>442, 449</td>
</tr>
<tr>
<td>transition matrix</td>
<td>438</td>
</tr>
<tr>
<td>Viterbi algorithm</td>
<td>443</td>
</tr>
<tr>
<td>math module</td>
<td>138, 635</td>
</tr>
<tr>
<td>mathematical operations</td>
<td>611</td>
</tr>
<tr>
<td>matplotlib module</td>
<td>141</td>
</tr>
<tr>
<td>matrix</td>
<td>145</td>
</tr>
<tr>
<td>convolution</td>
<td>374</td>
</tr>
<tr>
<td>determinant</td>
<td>302</td>
</tr>
<tr>
<td>multiplication</td>
<td>154, 157</td>
</tr>
<tr>
<td>rotation</td>
<td>155</td>
</tr>
<tr>
<td>transpose</td>
<td>153</td>
</tr>
<tr>
<td>mean value</td>
<td>459</td>
</tr>
<tr>
<td>median value</td>
<td>458</td>
</tr>
<tr>
<td>memoisation</td>
<td>175</td>
</tr>
<tr>
<td>methods</td>
<td>103</td>
</tr>
<tr>
<td>Metropolis-Hastings</td>
<td>550</td>
</tr>
<tr>
<td>microarrays</td>
<td></td>
</tr>
<tr>
<td>colour channels</td>
<td>334</td>
</tr>
<tr>
<td>data model</td>
<td>323</td>
</tr>
<tr>
<td>philosophy</td>
<td>317</td>
</tr>
<tr>
<td>mode value</td>
<td>458</td>
</tr>
<tr>
<td>modules</td>
<td>41, 179</td>
</tr>
<tr>
<td>importing</td>
<td>41</td>
</tr>
<tr>
<td>search path</td>
<td>41</td>
</tr>
<tr>
<td>test code block</td>
<td>61</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td></td>
</tr>
<tr>
<td>integration example</td>
<td>547</td>
</tr>
<tr>
<td>Metropolis-Hastings</td>
<td>550</td>
</tr>
<tr>
<td>minimisation</td>
<td>549</td>
</tr>
<tr>
<td>travelling salesman</td>
<td>554</td>
</tr>
<tr>
<td>multi-dimensional arrays</td>
<td>148</td>
</tr>
<tr>
<td>multiprocessing module</td>
<td>583</td>
</tr>
<tr>
<td>ndimage module</td>
<td>368</td>
</tr>
<tr>
<td>Needleman-Wunsch alignment</td>
<td>221</td>
</tr>
<tr>
<td>negative binomial distribution</td>
<td>437</td>
</tr>
<tr>
<td>neural network</td>
<td>518, 523</td>
</tr>
<tr>
<td>prediction</td>
<td>527</td>
</tr>
<tr>
<td>sequence example</td>
<td>531</td>
</tr>
<tr>
<td>training</td>
<td>525, 528</td>
</tr>
<tr>
<td>None value</td>
<td>25</td>
</tr>
<tr>
<td>normal distribution</td>
<td>467</td>
</tr>
<tr>
<td>normal tailed test</td>
<td>469</td>
</tr>
<tr>
<td>NOT operation</td>
<td>49</td>
</tr>
<tr>
<td>nucleotides</td>
<td>184</td>
</tr>
<tr>
<td>null hypothesis</td>
<td>456</td>
</tr>
<tr>
<td>null value</td>
<td></td>
</tr>
<tr>
<td>Needleman-Wunsch value</td>
<td></td>
</tr>
<tr>
<td>NumPy arrays</td>
<td>148</td>
</tr>
<tr>
<td>array() function</td>
<td>150, 646</td>
</tr>
<tr>
<td>comparing</td>
<td>178</td>
</tr>
<tr>
<td>creation</td>
<td>151</td>
</tr>
<tr>
<td>data types</td>
<td>151</td>
</tr>
<tr>
<td>element-wise operations</td>
<td>151</td>
</tr>
<tr>
<td>FFT</td>
<td>388</td>
</tr>
<tr>
<td>image pixmaps</td>
<td>366</td>
</tr>
<tr>
<td>inbuilt methods</td>
<td>153</td>
</tr>
</tbody>
</table>
## Index

NumPy arrays (cont.)  
indexing, slicing, 152  
mathematics functions, 152  
mean value, 459  
ndarray object, 150  
numeric, 148  
range, 153  
size, shape, 151  
sorting, 152  
standard deviation, 461  
variance, 461  
numpy module, 645  
numpy.linalg module, 301  
object orientation, 100  
objects  
attribute cardinality, 122  
attribute getters, setters, 132  
attributes, 105  
class information, 114  
custom classes, 102  
key identifier, 121  
parent-child links, 124  
types, 115  
one-tail test, 463  
open() function, 81  
optimisation  
Monte Carlo, 549  
SciPy, 396  
simulated annealing, 560  
optimising speed, 602  
OR operation, 49  
os module, 97, 636  
os.path module, 97, 638  
external product, 398  
parallel processing, 583  
passive substitutions, 275  
PDB files, 129  
downloading, 287  
Pearson’s chi-squared test, 476  
Pearson’s correlation coefficient, 481  
percent point function, 473–476  
performance, 162  
phylogenetic trees  
construction, 263  
neighbour-joining, 263  
philosophy, 251  
pickle module, 99, 640  
pie charts, 143  
plots, 141  
Poisson distribution, 435  
Poisson test, 467  
pow() function, 165  
principal component analysis, 504  
print statement, 20  
print() function, 20  
probability basics, 423  
Bayes’ theorem, 429  
binomial distribution, 431  
combining, 426  
conditional, 428  
geometric distribution, 436  
hidden Markov models, 442  
Markov chains, 438  
negative binomial distribution, 437  
Poisson distribution, 435  
posterior, 430  
probability density function, 467  
property() function, 135  
proteins  
backbone angles, 294  
estimating charge, 202  
hydrophobicity, 196  
isoelectric point, 200  
membranes, 281  
primary structure, 281  
quaternary structure, 281  
Ramachandran plot, 296  
residue similarity, 216  
secondary structure, 281  
structure, 279  
structure hierarchy, 281  
Python installation, 18  
pyqt module, 577  
Python 3  
dictionary views, 40  
division, 26  
print() function, 20  
rounding numbers, 141  
Qt graphics, 577  
Qt signals and slots, 578  
quantile normalisation, 332  
quitting a program, 61  
R statistical package, 668  
raise statement, 59  
Ramachandran plot, 296  
random module, 636  
range() function, 54  
regular expressions, 658  
code syntax, 665  
relative entropy, 198  
residues, 182  
return statement, 66  
reverse complement, 69  
RGB colour model, 362  
RGBA colour model, 362  
RNA, 185  
root-mean-square deviation, 302  
rotation matrix, 155  
round() function, 140

© in this web service Cambridge University Press
Index

rounding numbers, 140
RPy, 668
SAM files
creation with Bowtie, 353
reading, 357
scatter plots, 142
scipy.optimize module, 396
scipy.stats module, 431
scope of variables, 72
self variable, 103, 121
self-organising map, 518, 520
sequence collection types, 621
sequence complexity, 197
sequences
alignment, 208
alignment score, 218
annotation files, 358
consensus, 233
conservation, 253
conservation, substitution, 272
database fetching, 206
distance matrix, 265
genome data, 344
genome indexing, 349
genomic mapping, 351
HMM example, 449
identity score, 214
pairwise alignment, 221
profile, 235
quality scores, 352
reading from file, 205
similarity score, 215, 217
variation mechanisms, 244
vector representation, 532
writing to file, 206
setattr() function, 114
sets, 33
adding, removing items, 38
membership, 38
operations, 38, 170, 627
restrictions, 33
size, 38
shutil module, 641
signals
peak finding, 393
peak fitting, 395
plotting, 384
simulation, 384
spectrum peaks, 389
simulated annealing, 557
sin() function, 139
singular value decomposition, 301
slicing
lists, tuples, 35
NumPy arrays, 152
strings, 28
speed improvement, 162
splitting text strings, 29
SQL, 401
basics, 403
databases. See databases injection attack, 412
Python modules, 403
running, 404
sqrt() function, 138
standard normal distribution, 470
standard Python functions, 614
standard score, 470
statement, 23
statistical error, 456
statistics
binomial test, 464
categorical counts, 477
CDF, 466
chi-squared test, 476
certainty interval, 475
correlation, 481
covariance, 480
ear function, 471
G-test, 479
hypothesis testing, 462
likelihood ratio, 479
linear regression, 484
mean, 459
median, 458
mode, 458
normal distribution, 467
normal tailed test, 469
one-sample T-test, 474
one-tail test, 463
PDF, 467
Poisson test, 467
PPF, 476
probability intervals, 475
p-values, 463
skewness, 462
standard deviation, 461
standard error of mean, 462
survival function, 466
tail tests, 463
T-tests, 472
two-sample T-test, 474
two-tail test, 463
variance, 460
Welch’s T-test, 474
Z-score, 470
stopping a program, 61
string module, 634
strings
data type, 25
formatting, 31, 653
formatting, new-style, 654
joining, 167
methods, 622
operations, 621
quotation marks, 25
triple quotes, 25
structures, 286
affine transform, 292
atomic distances, 293
backbone alignment, 309
centre of mass, 288
centring, 300
coordinate alignment, 299, 301
coordinate arrays, 290
data model classes, 119
dihedral angle, 294
downloading, 287
ensemble alignment, 304
homologous alignment, 308
loading, 288
mirroring, 288
one-letter sequence, 309
PyMol visualisation, 314
RMSD, 302
rotation, 290–291
SQL database, 406
sub-selections, 297
torsion angle, 294
subclassing, 105, 115
constructor, 110
subprocess module, 226
substitution matrices, 215–216
calculating, 258
substitution rates, 272
substrings, 28
support vector machines, 534, 541
sys module, 640
sys.argv attribute, 84
text functions, 622
textual data. See Strings: data type
thumbnail images, 366
time module, 641
timing performance, 163
tkinter module, 570
torsion angle, 157
transcription, 185
translate() function, 70, 167
translation, of codons, 185, 190
travelling salesman problem, 552
trigonometric functions, 139
triple-quoted string, 25
True value, 47
try statement, 58
T-statistic, 472
T-tests, 472
tuples, 32
manipulation, 34
membership, 35
operations, 621
two-tailed test, 463
type checking, 177
type classes, 115
type() function, 25, 115
types module, 177
undefined value. See None value
urllib module, 287, 345
urlopen() function, 288, 345
variables, 23
as attributes, 105
naming, 23
naming convention, 160
scope, 72
swapping, 165
vector distance, 516
vector space, 144
Viterbi algorithm, 443
while loops, 52
whitespace
in Python code, 22
removal, 29
splitting on, 84
with statement, 83
XML files, 90, 229
zip() function, 164, 169
Z-score, 470