

Question 1:

report the frequencies for all amino acids in all secondary structures as computed from trainCounts.

AC	71295	0.313366709
AE	38096	0.167445377
AH	118122	0.519187915
CC	15219	0.40029985
CE	11012	0.289644651
CH	11788	0.310055499
DC	91108	0.54661083
DE	20452	0.122703656
DH	55118	0.330685513
EC	67859	0.343609011
EE	30453	0.154200993
EH	99177	0.502189995
FC	35731	0.305463654
FE	36291	0.310251084
FH	44951	0.384285262
GC	131513	0.665208243
GE	29389	0.148653023
GH	36800	0.186138734
HC	29394	0.446032685
HE	14397	0.21846406
HH	22110	0.335503255
KC	66098	0.391965938
KE	29944	0.177570093
KH	72590	0.430463969
LC	73825	0.267783626
LE	66075	0.239672239
LH	135789	0.492544135
MC	15908	0.325510016
ME	10376	0.212314051
MH	22587	0.462175933
NC	70299	0.569015339
NE	17040	0.137925452
NH	36206	0.293059209
PC	90246	0.701212121
PE	12963	0.100722611

PH	25491	0.198065268
QC	38836	0.350850566
QE	18797	0.169815071
QH	53058	0.479334363
RC	52829	0.354001099
RE	30235	0.202601284
RH	66170	0.443397617
SC	87485	0.499104311
SE	32805	0.18715342
SH	54994	0.31374227
TC	67458	0.436392571
TE	41901	0.271061773
TH	45222	0.292545656
VC	50802	0.254683638
VE	82384	0.413012418
VH	66285	0.332303944
WC	11710	0.298222381
WE	11523	0.293459991
WH	16033	0.408317628
XC	50033	0.25260134
XE	64830	0.327306875
XH	83208	0.420091785
YC	31418	0.309564395
YE	32141	0.316688179
YH	37932	0.373747426

Based on the frequencies above, we can decide in which secondary structure each amino acid should be classified as follows:

AH, CC, DC, EH, FH, GC, HC, KH, LH, MH, NC, PC, QH, RH, SC, TC, VE, WH, XH, YH

Question 2:

Now compute the accuracy (correctly classified amino acids/total amino acids) of a basic predictor based on the decisions above, both on the training set (rely on trainCounts) and on the test set (testCounts).

Accuracy(Training set)= 0.490575806

Accuracy(Test set)= 0.489358187

Question 3:

estimate the accuracy of at least two other baseline predictors: one in which all amino acids are predicted to be in a coil; one in which amino acids are assigned randomly to a secondary structure.

Accuracy(coil-Training set)=coil that correctly predicted / total aa=0.202067341

Accuracy(coil-Test set)= coil that correctly predicted / total aa=0.203357257

Accuracy(random)=1/3