

HMM to break a simple substitution ciphertext message, using 200 iterations of the Baum-Welch re-estimation algorithm

SHIFT =3

From the final B matrix, determined the ciphertext letters that correspond to consonants and vowels

Since my plaintext has space included I have denoted it by *.

PT: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z *

CT: d e f g h i j k l m n o p q r s t u v w x y z * a b c

Final value of pi matrix

0.0 1.0

Final value of b matrix

0.036776		0.000000	yA
0.000000		0.000000	z B
0.000000		0.352434	* C
0.008040		0.182117	a D
0.030969		0.000000	b E
0.030144		0.017470	c F
0.061939		0.000000	d G
0.000000		0.149266	e H
0.032905		0.000000	fI
0.063675		0.000213	gJ
0.096774		0.000005	hK
0.000000		0.130608	I L
0.000000		0.000000	jM
0.019356		0.000000	kN
0.073838		0.005914	l O
0.040647		0.000000	m P
0.121942		0.000000	n Q
0.000000		0.120242	o R
0.040641		0.000007	pS
0.001936		0.000000	qT
0.079359		0.000000	rU
0.086289		0.000870	s V
0.110046		0.014814	t W

0.008593		0.026040 u	X
0.009678		0.000000 v	Y
0.042583		0.000000 w	Z
0.003871		0.000000 x *	

So, the state 1 corresponds to vowel and white space (highlighted ones) and state 2 corresponds to consonants.

I tested all with caesars cipher since it is convenient to validate.

Caesars cipher (shift of 3)

Hence Plaintext to Cipher text mapping is:

Pt: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z *

CT: d e f g h i j k l m n o p q r s t u v w x y z * a b c

Since my plaintext has space included I have denoted it by *.

Length of observation_Sequence=1000

Restarts=500

Final value of b matrix

P=a-----> C=d----->0.9927955494132692

P=b-----> C=e----->0.7183487515788692

P=c-----> C=f----->0.5828768572911597

P=d-----> C=g----->0.8213678630288404

P=e-----> C=h----->0.8012835217752436

P=f-----> C=i----->0.7389033141341376

P=g-----> C=j----->1.0

P=h-----> C=k----->0.9079583914662778

P=i-----> C=l----->0.9998648803901087

P=j-----> C=e----->0.7868851252251595

P=k-----> C=n----->0.9114261089236971

P=l-----> C=o----->1.0

P=m-----> C=p----->0.9272160689077378

P=n-----> C=q----->0.9074093344215488

P=o-----> C=r----->0.8219909144612054

P=p-----> C=s----->0.8810987640871363

P=q-----> C=t----->0.8467481856591311

P=r-----> C=u----->0.8816658777850984

P=s-----> C=v----->0.7777555475021002
P=t-----> C=w----->0.7690977114753413
P=u-----> C=x----->0.8398643382273491
P=v-----> C=y----->0.5394133780283057
P=w-----> C=z----->0.8218218137103867
P=x-----> C=u----->0.7242472875704583
P=y-----> C=a----->0.871521928386287
P=z-----> C=p----->0.6846940953880006
P=*-----> C=c----->0.9999999573519239 c

So apart from the highlighted ones, rest plaintext ciphertext mappings are predicted correctly.

Hence accuracy=24/27