Naïve Bayes Classifer Training Data

John	is	studying	with	Mary
noun	aux	verb	prep	noun

Mary	does	not	like	studying
noun	aux	adv	verb	noun

Susan	is	also	like	Mary
noun	verb	adv	verb	noun

W_i	John	is	studying	with	Mary	does	not	like	Susan	also
noun	1/6		1/6		3/6				1/6	
aux		1/2				1/2				
verb		1/4	1/4					2/4		
prep				1						
adv							1/2			1/2

T_{i-1}	*null*	noun	aux	verb	prep	adv
noun	3/6			2/6	1/6	
aux		1				
verb		1/4	1/4			2/4
prep				1		
adv			1/2	1/2		

	P(t)
noun	6/15
aux	2/15
verb	4/15
prep	1/15
adv	2/15

	John	is	studying	with	Mary	does	not	like	Susan	also
noun	0.17		0.17		0.50				0.17	
aux		0.50				0.50				
verb		0.25	0.25					0.50		
prep				1.00						
adv							0.50			0.50

T_{i-1}	*null*	noun	aux	verb	prep	adv
noun	0.50			0.33	0.17	
aux		1.00				
verb		0.25	0.25			0.50
prep				1.00		
adv			0.50	0.50		

	P(t)
noun	0.40
aux	0.13
verb	0.27
prep	0.07
adv	0.13

Constant	0.1

Tag Sequence

	John	did	not	like	playing	with	Tom
	noun	aux	adv	verb	noun	aux	verb/adv
noun	**0.03333**	0.00400	0.00400	0.00400	**0.01333**	0.00400	0.00400
aux	0.00133	**0.01333**	0.00133	0.00133	0.00133	**0.01333**	0.00133
verb	0.00267	0.00667	0.00667	**0.06667**	0.00267	0.00667	**0.00667**
prep	0.00067	0.00067	0.00067	0.00067	0.00667	0.00667	0.00067
adv	0.00133	0.00133	**0.03333**	0.00133	0.00667	0.00133	**0.00667**

Notice how Tom has an equal chance of being a “verb” or “adv”. This isn’t correct but is the best our model can come up with. Based off the training data our model used to learn from and calculate probabilities, its reasonable for new words to trump the system. We need a larger training data for more accurate decoding.

.2666*0.25*Constant == .1333*0.5*Constant
Therefore Tom could be tagged either verb or adv