Part 1 - POS Tagging (Total 10 Marks)

Consider the following HMM with three possible observations "snow", "fell", "storm" and three possible Part-of-Speech (POS) tags ("N", V", "J"):

State transition probabilities (A) – e.g., $a_{N,J} = P(s_{i+1} = J|s_i = N) = 0.1$

A	N (noun)	V (verb)	J (adj)
N	0.4	0.5	0.1
V	0.5	0.1	0.4
J	0.5	0.1	0.4

Emission probabilities (B) – e.g., $b_N(snow) = P(o_i = snow|s_i = N) = 0.4$

B	snow	fell	storm
N	0.4	0.2	0.4
V	0.3	0.5	0.2
J	0.2	0.4	0.4

Initial state distributions (π) – e.g., $\pi[J] = P(s_1 = J|s_0 =< S >) = 0.3$

	N	V	J
π	0.4	0.3	0.3

Question 1.1. Draw the forward trellis diagram of the sentence "snow fell" using the given HMM. Clearly show the forward arrows to illustrate the computation of the forward algorithm. (5 marks)

Question 1.2. Find the most likely state sequence for the sentence "snow fell" using the Viterbi algorithm (show your steps clearly). What is the joint probability of the sentence "snow fell" and its most likely state sequence? (5 marks)

1.1) o1 = snow o2 = fell

		o1 = snow	o2 = fell
(s1 = N)		α1(1) = 0.16	α2(1) = 0.0278
(s2 = V)	π	α1(2) = 0.09	α2(2) = 0.0475
(s3 = J)		α1(3) = 0.06	α2(3) = 0.0304

α1(1) = P(N|π)*P(o1|N) = 0.4*0.4 = 0.16,

α1(2) = P(V|π)*P(o1|V) = 0.3*0.3 = 0.09,

α1(3) = P(J|π)*P(o1|J) = 0.3*0.2 = 0.06,

α1(1)* P(N|N) = 0.16 * 0.4 = 0.064, α1(1)* P(V|N) = 0.16 * 0.5 = 0.08, α1(1)* P(J|N) = 0.16 * 0.1 = 0.016

α1(2)* P(N|V) = 0.09 * 0.5 = 0.045, α1(2)* P(V|V) = 0.09 * 0.1 = 0.009, α1(2)* P(J|V) = 0.09 * 0.4 = 0.036

α1(3)* P(N|J) = 0.06 * 0.5 = 0.03, α1(3)* P(V|J) = 0.06 * 0.1 = 0.006, α1(3)* P(J|J) = 0.06 * 0.4 = 0.024

P(o2|N) = 0.2, P(o2|V) = 0.5, P(o2|J) = 0.4

α2(1) = (α1(1) * P(N|N) + α1(2) * P(N|V) + α1(3) * P(N|J))*P(o2|N) = (0.064+0.045+0.03)*0.2 = 0.0278

α2(2) = (α1(1) * P(V|N) + α1(2) * P(V|V) + α1(3) * P(V|J))*P(o2|V) = (0.08+0.009+0.006)*0.5 = 0.0475

$\alpha2(3) = (\alpha1(1) * P(J|N) + \alpha1(2) * P(J|V) + \alpha1(3) * P(J|J))*P(o2|J) = (0.016+0.036+0.024)*0.4 = 0.0304$

1.2)

Given:

$P(o2|N) = 0.2$, $P(o2|V) = 0.5$, $P(o2|J) = 0.4$

Step 1:

$v1(1) = P(N|\pi)*P(o1|N) = 0.4*0.4 = 0.16$

$v1(2) = P(V|\pi)*P(o1|V) = 0.3*0.3 = 0.09$

$v1(3) = P(J|\pi)*P(o1|J) = 0.3*0.2 = 0.06$

Step 2:

$v2(1) = \max(v1(1) * p(N|N)* P(o2|N), v1(2) * p(N|V)* P(o2|N), v1(3) * p(N|J)* P(o2|N)) = \max(0.16*0.4*0.2, 0.09*0.5*0.2, 0.06*0.5*0.2) = \max(0.0128,0.009,0.006)$

= **0.0128 (N->N)**

$v2(2) = \max(v1(1) * p(V|N)* P(o2|V), v1(2) * p(V|V)* P(o2|V), v1(3) * p(V|J)* P(o2|V)) =$

$\max(0.16*0.5*0.5, 0.09*0.1*0.5, 0.06*0.1*0.5) = \max(0.04,0.0045,0.003)$

= **0.04 (N->V)**

$v2(3) = \max(v1(1) * p(J|N)* P(o2|J), v1(2) * p(J|V)* P(o2|J), v1(3) * p(J|J)* P(o2|J)) =$

$\max(0.16*0.1*0.4, 0.09*0.4*0.4, 0.06*0.4*0.4) = \max(0.0064,0.0144,0.0096)$

= **0.0144 (V->J)**

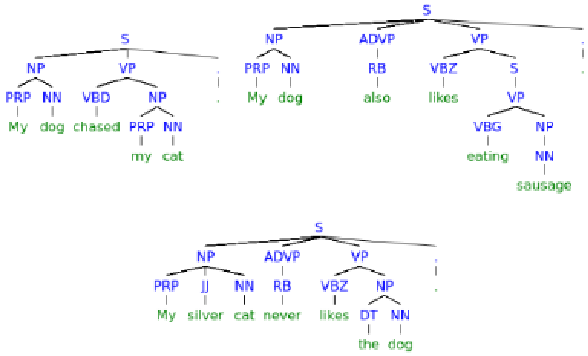
Step 3:

P* = max(N->N,N->V,N->J) = max(0.0128,0.04,0.0144) = **0.04 (N->V)**

As N->V has the highest joint probability of 0.04, (π,N,V) is the most likely state sequence using the Viterbi Scores for the phrase ‘snow fell’

Part 2 - PCFG (Total 10 Marks)

Consider the following parse trees:



Question 2.1. Convert the grammar rules into the CNF format and list them. Then estimate their corresponding probabilities (based on all trees) using Maximum Likelihood estimation. (5 Marks)

Note. Structure you response as one rule per line:
[0.2] A $\rightarrow$ B C
[0.1] D $\rightarrow$ E
...
where A $\rightarrow$ B C is the grammar rule, and [0.2] is its probability.

Question 2.2. Using the estimated probabilities and the CKY parsing algorithm, calculate the probability of the following sentence (and report the final CKY chart): (5 Marks)

`The dog chased my silver cat .`

2.1)

[0.25] S -> NP X3

[1] X3 -> VP .

[1] . -> .

[0.5] S -> X1 VP

[1] X1 -> NP ADVP

[0.0625] S -> VBD NP

[0.0625] S -> VBZ S

[0.0625] S -> VBG NP

[0.0625] S -> VBZ NP

[0.5] NP -> PRP NN

[0.02822] NP -> sausage

[0.083] NP -> dog

[0.05478] NP -> cat

[0.166] NP -> DT NN

[0.166] NP -> X2 NN

[1] X2 -> PRP JJ

[0.25] VP -> VBD NP

[0.25] VP -> VBZ S

[0.25] VP -> VBG NP

[0.25] VP -> VBZ NP

[0.5] ADVP -> Also

[0.5] ADVP -> Never

[0.5] RB -> Also

[0.5] RB -> Never

[0.5] NN -> dog

[0.33] NN -> cat

[0.17] NN -> sausage

[1] PRP -> My

[1] VBD -> chased

[1] VBZ -> likes

[1] VBG -> eating

[1] DT -> the

[1] JJ -> silver

2.2)

The	dog	chased	my	silver	cat	.
DT: 1	NP: 0.166 * 1* 0.5 = 0.083	None	None	None	NP: 0.166 * 1 * 0.33 = 0.05478	S: 0.25 * 0.083 * 0.013675= **0.00028375625**
	NN:0.5, NP: 0.083	None	None	None	None	S: 0.25 * 0.083 * 0.013675= **0.00028375625**
		VBD: 1	None	none	VP: 0.25 * 1 * 0.0547 = 0.013675 S: 0.0625 *1 * 0.0547 = 0.00341875	X3: 1 * 1 * 0.013675 = **0.013675**
			PRP: 1	X2 : 1*1*1 = 1	NP: 0.166 * 1 * 0.33 = **0.0547** NP: 0.5 * 1 * 0.33 = 0.165	
				JJ: 1		
					NN: **0.33**, NP: 0.0547	
						.: **1**

The sentence "The dog chased my silver cat." has a probability of **0.00028375625** with a parse tree of: S-> NP VP, where NP -> DT(The) NN(Dog), VP-> VBD(chased)NP, NP -> X2 NN(cat), X2 -> PRP(my) JJ(silver).