KUSHAGRA SETH
2005986
kuseth@ucsc.edu

Homework Assignment 3
NLP-201

Prof: Jeff Flanigan

1 Deriving the Viterbi Algorithm

1.1

$$P(t|w) = \frac{P(t,w)}{P(w)} \tag{1}$$

$$P(t|w) \; \alpha \; P(t,w) \tag{2}$$

Taking log on both sides,

$$log(P(t|w)) \; \alpha \; log(P(t,w)) \tag{3}$$

Since log is a monotonic function the best *'t'* which maximizes $P(t|w)$ will always be the best *'t'* which maximizes $log(P(t|w))$.

Hence,

$$argmax \; P(t|w) = argmax \; log(P(t|w)) \tag{4}$$

1.2

We have,

$$\pi_j(t_j) = argmax_{t_1 \ldots . t_{n+1}} \sum_{i=1}^{j} score(w, i, t_i, t_{i-1}) \tag{5}$$

RHS of (5) can be written as:

$$argmax_{t_1 \ldots t_{n+1}}[(\sum_{i=1}^{j-1} score(w, i-1, t_{i-1}, t_{i-2}) + score(w, i, t_i, t_{i-1})] \tag{6}$$

$$argmax_{t_1 \ldots t_{n+1}}[(\sum_{i=1}^{j-1} score(w, i-1, t_{i-1}, t_{i-2}) + score(w, i, t_i, t_{i-1})] \tag{7}$$

Simplifying (6) further we have,

$$argmax_{t_1 \ldots t_{n+1}}[(\sum_{i=1}^{j-1} score(w, i-1, t_{i-1}, t_{i-2}) + score(w, i, t_i, t_{i-1})] \tag{8}$$

Hence Proved,

$\pi_j(t_j)$ can be written as a compound of π_{j-1} state.

1.3

To compute $\hat{t}$, we must use Viterbi Algorithm. In the Viterbi Algorithm, there are *K* tags and *N* positions in the sequence. Subsequently, then there are *N X K* Viterbi variables that we have to compute.

Finding the maximum over K possible preceding tags is necessary for computing each variable. As a result, the *Total Time Complexity* of populating the Viterbi is $O(NK^2)$, plus an additional factor for the number of active features at each point. We trace backwards the pointers to the start of the sequence after the Viterbi is completed, which requires O(M) operations.

1.4

{placeholder}

2 Programming: Hidden Markov Model

2.1 Baseline Tagger Results

We have implemented a Baseline HMM Tagger with Add-α smoothing. We have added $\langle START \rangle$ and $\langle STOP \rangle$ tokens at the beginning and end of gold sequences.

S.No	Data Set	Accuracy
1.	Dev Set	0.91
2.	Test Set	0.92

Tabelle 1: Baseline Tagger Scores on Dev and Test Set

3 Programming: Viterbi Decoding

3.1 HMM POS Tagger Results

We have implemented Viterbi decoding and α-Smoothing in the HMM POS Tagger, with $\alpha = 1$ and calculated to find the best tag sequence on the test dataset.

S.No	Data Set	Accuracy
1.	Dev Set	0.84
2.	Test Set	0.85

Tabelle 2: HMM POS Tagger Scores on Dev and Test Set with $\alpha = 1$

	#	$	''	(	)	,	.	:	CC	CD	DT	EX	FW	IN	IN|RB	JJ	JJR	JJR|RBR	JJS	JJ|IN	JJ|JJR	JJ|NN	JJ|RB
#	<22>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
$	.	<1136>	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
''	.	.	<1396>	.	.	.	.	.	.	.	.	.	24	.	.	.	.	.	.	.	.	.	.
(	.	.	.	<236>	.	.	.	.	.	.	.	.	13	.	.	.	.	.	.	.	.	.	.
)	.	.	.	.	<237>	.	.	.	.	.	.	.	14	.	.	.	.	.	.	.	.	.	.
,	.	.	.	.	.	<8950>	.	.	.	.	.	.	106	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	<7030>	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.
:	.	.	.	.	.	.	.	<882>	.	.	.	.	99	.	.	.	.	.	.	.	.	.	.
CC	.	.	.	.	1	.	.	.	<3936>	.	9	.	333	.	.	.	.	.	.	.	.	.	.
CD	11	65	5	1	1	1	2	1	1	<5008>	96	4	416	6	.	63	5	.	3	.	1	.	1
DT	.	.	14	.	5	.	.	1	1	.	<14392>	37	284	38	.	3	.	.	.	.	.	.	1
EX	.	.	.	.	.	.	.	.	.	.	.	<172>	1	.	.	.	.	.	.	.	.	.	.
FW	.	.	2	.	1	3	.	.	.	.	2	.	<14>	.	.	2	.	.	.	.	.	.	.
IN	3	8	7	2	.	.	.	.	1	.	42	.	1025	<16516>	10	10	.	.	.	.	.	.	7
IN|RB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.	.	.	.	.	.	.	.
JJ	9	18	13	1	10	6	5	3	4	44	158	7	1658	21	23	<7571>	7	.	7	.	3	1	10
JJR	.	.	1	.	.	.	.	.	.	1	7	2	52	.	.	5	<452>	.	.	.	.	.	.
JJR|RBR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	<.>	.	.	.	.	.
JJS	.	.	.	.	.	.	.	.	.	.	2	.	9	.	.	10	.	.	<330>	.	.	.	.
JJ|IN	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	<.>	.	.	.
JJ|JJR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.	.
JJ|NN	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.
JJ|RB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>

Abbildung 1: Confusion Matrix with $\alpha = 1$

4 Evaluation

4.1 Hyper-Parameter Tuning on Dev Set

We have experimented with different values of the hyper-parameter α to see if we can get a better accuracy score than in the standard α-smoothing where $\alpha = 1$.

The following Table presents an Ablation Study of Accuracy, Macro-F1, Precision & Recall, and Weighted-F1, Precision Recall for different values of α.

	alpha	accuracy	macro_precision	macro_recall	macro_f1	weighted_precision	weighted_recall	weighted_f1
0	1.000000e+00	0.83797	0.50493	0.58993	0.50834	0.93696	0.83797	0.88052
1	1.000000e-01	0.92474	0.46017	0.52691	0.47118	0.94865	0.92474	0.93574
2	1.000000e-02	0.94198	0.48299	0.51756	0.49066	0.95096	0.94198	0.94611
3	1.000000e-03	0.94878	0.50768	0.52531	0.51285	0.95042	0.94878	0.94933
4	1.000000e-04	0.94899	0.61562	0.63055	0.61964	0.95033	0.94899	0.94940
5	1.000000e-05	0.94900	0.65940	0.67190	0.66216	0.95028	0.94900	0.94938
6	1.000000e-06	0.94898	0.65940	0.67188	0.66215	0.95026	0.94898	0.94936
7	1.000000e-07	0.94898	0.65940	0.67188	0.66215	0.95026	0.94898	0.94936
8	1.000000e-08	0.94898	0.65940	0.67188	0.66215	0.95026	0.94898	0.94936
9	1.000000e-09	0.94898	0.65940	0.67188	0.66215	0.95026	0.94898	0.94936

4.2 HMM Evaluation on Test Set with α = 1e-5

We got the highest accuracy with α = 1e-5 and have reported the accuracy on Dev and Test Set. We have also plotted the Confusion Matrix as required for the best result on the Test Set, which we obtained with α-Smoothing and α = 1e-5

S.No	Data Set	Accuracy
1.	Dev Set	0.95
2.	Test Set	0.95

Tabelle 3: HMM POS Tagger Scores on Dev and Test Set with α = 1e-5

	#	$	''	(	)	,	.	:	CC	CD	DT	EX	FW	IN	IN|RB	JJ	JJR	JJR|RBR	JJS	JJ|IN	JJ|JJR	JJ|NN	JJ|RB
#	<22>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
$	.	<1136>	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
''	.	.	<1396>	.	.	.	.	.	.	.	.	.	24	.	.	.	.	.	.	.	.	.	.
(	.	.	.	<236>	.	.	.	.	.	.	.	.	13	.	.	.	.	.	.	.	.	.	.
)	.	.	.	.	<237>	.	.	.	.	.	.	.	14	.	.	.	.	.	.	.	.	.	.
,	.	.	.	.	.	<8950>	.	.	.	.	.	.	106	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	<7030>	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.
:	.	.	.	.	.	.	.	<882>	.	.	.	.	99	.	.	.	.	.	.	.	.	.	.
CC	.	.	.	.	1	.	.	.	<3936>	.	9	.	333	.	.	.	.	.	.	.	.	.	.
CD	11	65	5	1	1	1	2	1	1	<5008>	96	4	416	6	.	63	5	.	3	.	1	.	1
DT	.	.	14	.	5	.	.	1	1	.	<14392>	37	284	38	.	3	.	.	.	.	.	.	1
EX	.	.	.	.	.	.	.	.	.	.	.	<172>	1	.	.	.	.	.	.	.	.	.	.
FW	.	.	2	.	1	3	.	.	.	.	2	.	<14>	.	.	2	.	.	.	.	.	.	.
IN	3	8	7	2	.	.	.	.	1	.	42	.	1025	<16516>	10	10	.	.	.	.	.	.	7
IN|RB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.	.	.	.	.	.	.	.
JJ	9	18	13	1	10	6	5	3	4	44	158	7	1658	21	23	<7571>	7	.	7	.	3	1	10
JJR	.	.	1	.	.	.	.	.	.	1	7	2	52	.	.	5	<452>	.	.	.	.	.	.
JJR|RBR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	<.>	.	.	.	.	.
JJS	.	.	.	.	.	.	.	.	.	.	2	.	9	.	.	10	.	.	<330>	.	.	.	.
JJ|IN	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	<.>	.	.	.
JJ|JJR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.	.
JJ|NN	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>	.
JJ|RB	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	<.>

Abbildung 2: Confusion Matrix with α = 1e-5