

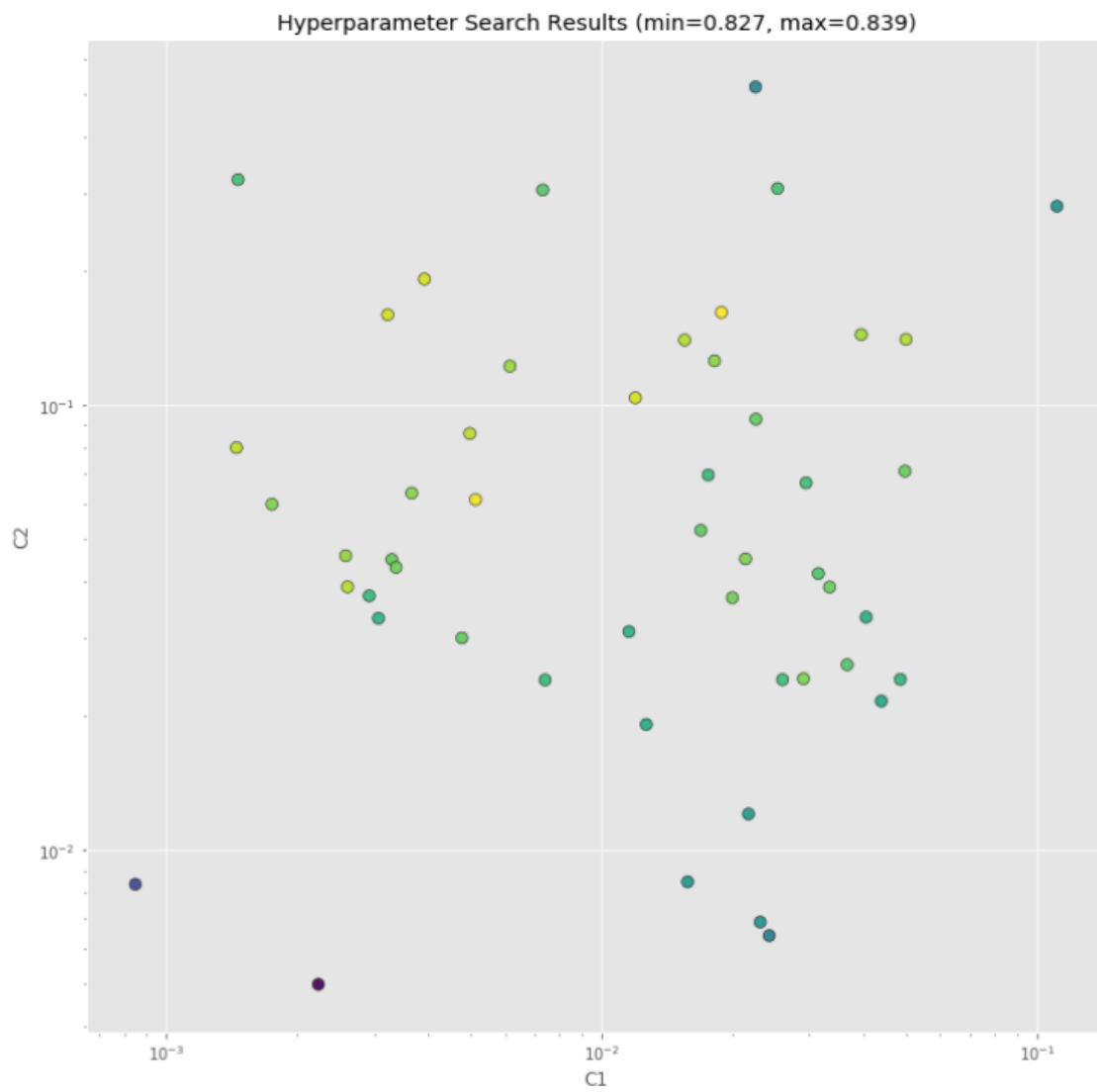
18MA20015, Harshal Dupare

Assignment 4

Features Used

number	feature	details
1	bias	1 for all words
2	index	location at which this word comes in the sentence
3	caps_freq	number of times a capital letter is there is the word divide by the length of word
4	word.lower()	word converted to lower case
5	word[-3:]	3 letter suffix of the word
6	word[-2:]	2 letter suffix of the word
7	word[-1:]	1 letter suffix of the word
8	word[:1]	1 letter prefix of the word
9	word[:2]	2 letter prefix of the word
10	word[:3]	3 letter prefix of the word
11	word.isupper()	1 if all letters are upper case else 0
12	word.istitle()	1 only if 1st letter is upper and all else lower if not then 0
13	word.isdigit()	1 if it is a digit else 0
14	word.length	length of the word
15	BOS	1 if its start of the sentence
16	EOS	1 if its end of the sentence

Hyperparameter tuning using Random Search



Best parameters are

$$c_1 = 0.0188$$

$$c_2 = 0.1616$$

Results on Train Data

	precision	recall	f1-score	support
X	1.000	1.000	1.000	2
PART	0.994	1.000	0.997	163
CCONJ	0.987	1.000	0.993	150
SCONJ	0.984	1.000	0.992	61
ADJ	0.982	0.979	0.981	569
ADP	0.983	0.992	0.987	1384
ADV	0.890	0.882	0.886	110
VERB	0.990	0.956	0.973	639
DET	0.937	0.974	0.955	230
COMMA	1.000	1.000	1.000	114
NOUN	0.980	0.984	0.982	1596
PRON	0.988	0.958	0.973	430
PROPN	0.987	0.967	0.977	707
NUM	0.980	0.987	0.984	152
PUNCT	1.000	1.000	1.000	563
AUX	0.971	0.996	0.983	728
accuracy			0.981	7598
macro avg	0.978	0.980	0.979	7598
weighted avg	0.981	0.981	0.981	7598

Accuracy Train = 0.9813108712819163 = 98.1%

Results on Test Data

	precision	recall	f1-score	support
X	0.000	0.000	0.000	0
PART	0.971	1.000	0.985	33
CCONJ	1.000	1.000	1.000	25
SCONJ	0.667	0.667	0.667	3
ADJ	0.704	0.734	0.719	94
ADP	0.967	0.974	0.970	303
ADV	0.600	0.429	0.500	21
VERB	0.866	0.848	0.857	99
DET	0.800	0.889	0.842	36
COMMA	0.000	0.000	0.000	0
NOUN	0.773	0.864	0.816	324
PRON	0.917	0.846	0.880	65
PROPN	0.633	0.528	0.576	144
NUM	0.815	0.880	0.846	25
PUNCT	1.000	0.836	0.911	134
AUX	0.911	0.964	0.937	138
micro avg	0.850	0.850	0.850	1444
macro avg	0.726	0.716	0.719	1444
weighted avg	0.850	0.850	0.848	1444

Accuracy Test = 0.8497229916897507 = 84.9%

10 most common Transition Features

VERB	->	AUX	3.225284
AUX	->	AUX	2.223406
DET	->	NOUN	2.041918
PART	->	NUM	2.013738
PROPN	->	PROPN	1.896575
NUM	->	NOUN	1.852065
ADJ	->	NOUN	1.652541
DET	->	PART	1.529679
ADP	->	ADP	1.490440
COMMA	->	CCONJ	1.484503

10 least common Transition Features

DET	->	PROPN	-1.498124
PROPN	->	ADJ	-1.529382
ADP	->	COMMA	-1.708934
CCONJ	->	AUX	-1.785534
PROPN	->	DET	-1.873113
ADV	->	AUX	-1.929062
PROPN	->	AUX	-2.094518
DET	->	ADP	-2.101813
ADJ	->	PRON	-2.559492
ADJ	->	ADP	-2.958521

