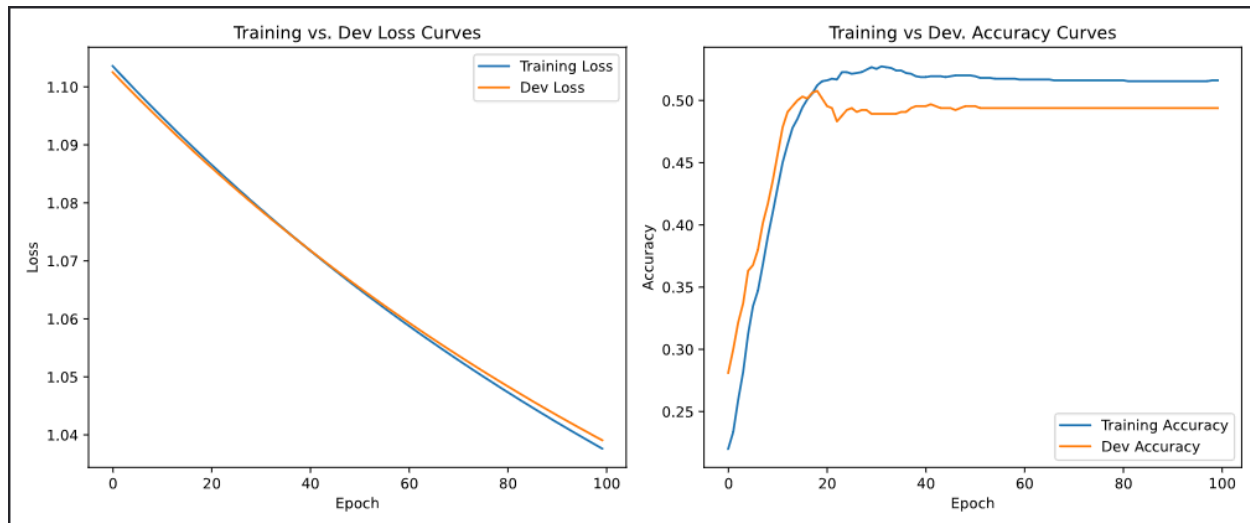Checkpoint 2.1:
Sum of first & last 5 feature vectors of X: [2. 4. 0. ... 0. 0. 0.]

Checkpoint 2.2:
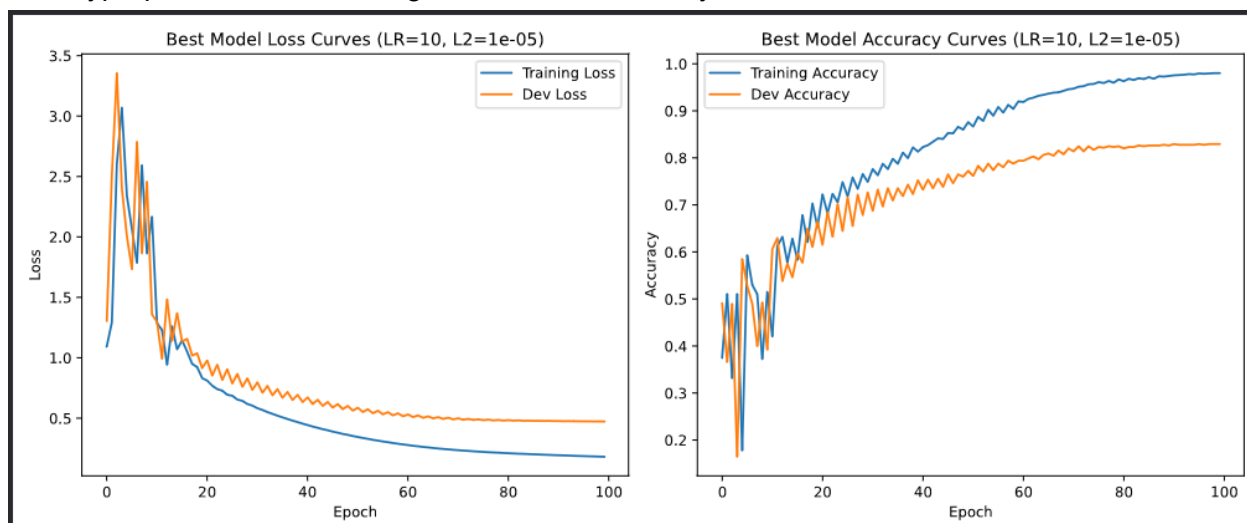

Checkpoint 2.3:
Dev set accuracy for different hyperparameters:

LR\L2	1e-5	1e-3	1e-1
0.1	0.5815	0.5938	0.5969
1	0.7523	0.7446	0.5969
10	0.8308	0.7538	0.3662

Best hyperparameters: Learning Rate = 10, L2 Penalty = 1e-05

Sample sentence : The horse raced past the barn fell.

Token	Predicted POS
The	O

Token	Predicted POS
horse	O
raced	O
past	O
the	O
barn	O
fell	O
.	O

Sample sentence : For 3 years, we attended S.B.U. in the CS program.

Token	Predicted POS
For	O
3	O
years	O
,	O
we	O
attended	O
S	O
.	O
B	O
.	O
U.	O
in	O
the	O
CS	O
program	O
.	O

Sample sentence : Did you hear Sam tell me to "chill out" yesterday? #rude

Token	Predicted POS
Did	O
you	O
hear	O
Sam	O
tell	O
me	O
to	O
"	O
chill	O
out	O
"	O
yesterday	O
?	O

#rude   O