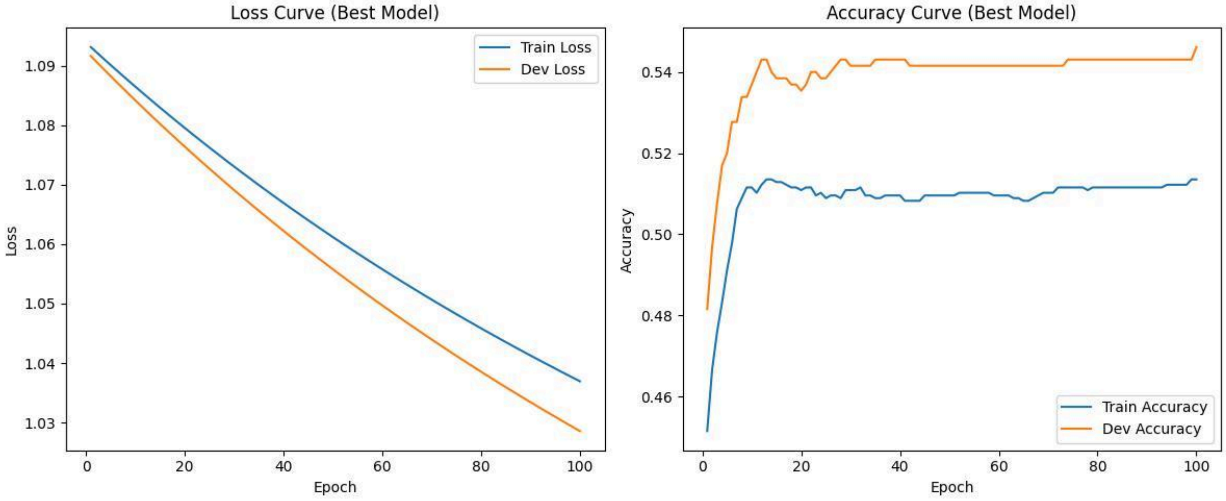
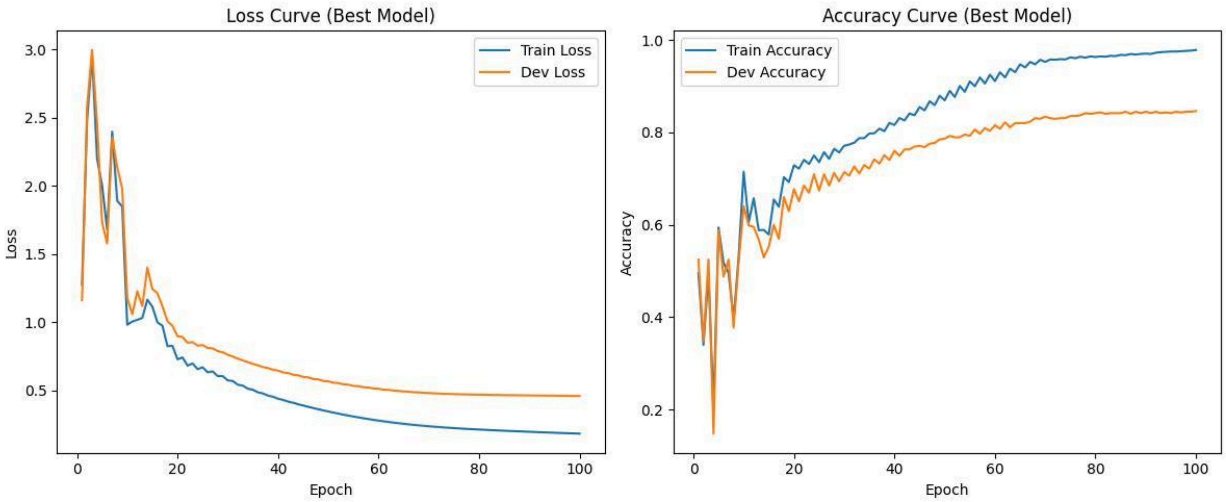
```
 ######### CHECKPOINT 2.1 #########
Sum of first 5 feature vectors:
 [2. 0. 0. ... 0. 0. 0.]
Sum of last 5 feature vectors:
 [0. 0. 0. ... 0. 0. 0.]
```

```
 ######### CHECKPOINT 2.2 #########

Training a Logistic Regression Model…
```



```
 ######### CHECKPOINT 2.3 #########

Dev Set Accuracy Table (LR rows, L2 columns):
LR\L2  1e-05   1e-03   1e-01
0.1    0.5815 0.5815 0.5815
1      0.7200 0.7169 0.5815
10     0.7969 0.7569 0.3754

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



```
######### CHECKPOINT 2.4 #########

Sentence: The horse raced past the barn fell.
Tokens: ['The', 'horse', 'raced', 'past', 'the', 'barn', 'fell', '.']
Predicted POS Tags:
The: O
horse: O
raced: N
past: V
the: O
barn: N
fell: O
.: O

Sentence: For 3 years, we attended S.B.U. in the CS program.
Tokens: ['For', '3', 'years', ',', 'we', 'attended', 'S.B.U.', 'in', 'the', 'CS', 'program', '.']
Predicted POS Tags:
For: N
3: O
years: N
,: O
we: O
attended: V
S.B.U.: N
in: O
the: O
CS: N
program: N
```

```
.: O

Sentence: Did you hear Sam tell me to "chill out" yesterday? #rude
Tokens: ['Did', 'you', 'hear', 'Sam', 'tell', 'me', 'to', 'chill', 'out', 'yesterday', '?', '#rude']
Predicted POS Tags:
Did: O
you: N
hear: O
Sam: N
tell: V
me: N
to: O
chill: V
out: O
yesterday: N
?: O
#rude: O
```

OBSERVATIONS:
The training dataset has only 333 Verbs in over 2000 training data samples.
So the model does not really learn to identify Verbs very well. Due to this, ‘raced’ is not identified as a Verb in the first sentence.
But none of the Nouns were shown to be wrong, as Nouns data was sufficient to identify POS in the example sentences. (there were false Positives for Nouns, but no false negatives).
In many cases, where context ambiguity is observed, the model tends to mistake the POS.

Next, more training data would cover more instances of a single word in multiple contexts. The current data is not enough for the model to learn all possible senses of a word. Along with this, it is also felt that the context window size is currently two (one word to the left and one word to the right). Just two words are again not sufficient to identify the POS of a word, given that we already have less data.

Also, the features are mostly one-hot encoded (first letter, adjacent words), and the adjacent words are used. But the word meanings are never used. Using the word meanings will increase the accuracy. Better approaches at word sense disambiguation such as Contextual Embeddings will also help.

Conclusion: the model does well on Nouns (‘N’), and suffers while identifying ‘O’ and ‘V’. This can be improved by using larger context windows, more training data (with more senses of the same words). The model can also be likely improved by using richer features such as contextual embeddings instead of just surface-level features being used right now (like capitalization, length of word etc).
These above improvements would likely get better qualitative (and quantitative) results.