Therefore,

Similarly:

1. Using Q3 and Q4 conclusion, we can get:
2. Not sure,

The CNN model outperform the bag of words, because we have front and back propagation process.

The output performance is:

Epoch: 0 Train accuracy: 0.657 Dev accuracy: 0.622

Epoch: 1 Train accuracy: 0.786 Dev accuracy: 0.679

Epoch: 2 Train accuracy: 0.837 Dev accuracy: 0.693

Epoch: 3 Train accuracy: 0.931 Dev accuracy: 0.720

Epoch: 4 Train accuracy: 0.956 Dev accuracy: 0.711

Epoch: 5 Train accuracy: 0.971 Dev accuracy: 0.722

Epoch: 6 Train accuracy: 0.983 Dev accuracy: 0.735

Epoch: 7 Train accuracy: 0.991 Dev accuracy: 0.737

Epoch: 8 Train accuracy: 0.998 Dev accuracy: 0.738

Epoch: 9 Train accuracy: 0.999 Dev accuracy: 0.739

Thanks to Qiushi’s help, I was able to debug my code. I was using one hot encoding to do the calculation and it is very computational heavy. Then I take his suggestions to use a fake numeral encoding to represent the one hot encoding and finish the assignment.