RESULTS FINAL LAB

Note: The results reported in this document are not deterministic due to a random initialization of parameters.

Using reviews with average length:

```
Epoch [1/6], Step [200/700], Loss: 0.6749
Epoch [1/6], Step [400/700], Loss: 0.6839
Epoch [1/6], Step [600/700], Loss: 0.6859
Epoch [1/6], Step [700/700], Loss: 0.6850
Epoch [2/6], Step [200/700], Loss: 0.6939
Epoch [2/6], Step [400/700], Loss: 0.6936
Epoch [2/6], Step [600/700], Loss: 0.6934
Epoch [2/6], Step [700/700], Loss: 0.6899
Epoch [3/6], Step [200/700], Loss: 0.5181
Epoch [3/6], Step [400/700], Loss: 0.4675
Epoch [3/6], Step [600/700], Loss: 0.4339
Epoch [3/6], Step [700/700], Loss: 0.4217
Epoch [4/6], Step [200/700], Loss: 0.2443
Epoch [4/6], Step [400/700], Loss: 0.2366
Epoch [4/6], Step [600/700], Loss: 0.2326
Epoch [4/6], Step [700/700], Loss: 0.2329
Epoch [5/6], Step [200/700], Loss: 0.1291
Epoch [5/6], Step [400/700], Loss: 0.1338
Epoch [5/6], Step [600/700], Loss: 0.1358
Epoch [5/6], Step [700/700], Loss: 0.1352
Epoch [6/6], Step [200/700], Loss: 0.0822
Epoch [6/6], Step [400/700], Loss: 0.0792
Epoch [6/6], Step [600/700], Loss: 0.0790
Epoch [6/6], Step [700/700], Loss: 0.0784
Test Accuracy of the model: 84.81333333333333 %
```

Encoding reviews as a list sorted digits with length of the longest review:

```
Epoch [1/6], Step [200/700], Loss: 0.6919
Epoch [1/6], Step [400/700], Loss: 0.6933
Epoch [1/6], Step [600/700], Loss: 0.6740
Epoch [1/6], Step [700/700], Loss: 0.6468
Epoch [2/6], Step [200/700], Loss: 0.3760
Epoch [2/6], Step [400/700], Loss: 0.3459
Epoch [2/6], Step [600/700], Loss: 0.3291
Epoch [2/6], Step [700/700], Loss: 0.3206
Epoch [3/6], Step [200/700], Loss: 0.1777
Epoch [3/6], Step [400/700], Loss: 0.1708
```

```
Epoch [3/6], Step [600/700], Loss: 0.1705
Epoch [3/6], Step [700/700], Loss: 0.1700
Epoch [4/6], Step [200/700], Loss: 0.0780
Epoch [4/6], Step [400/700], Loss: 0.0765
Epoch [4/6], Step [600/700], Loss: 0.0788
Epoch [4/6], Step [700/700], Loss: 0.0794
Epoch [5/6], Step [200/700], Loss: 0.0271
Epoch [5/6], Step [400/700], Loss: 0.0323
Epoch [5/6], Step [600/700], Loss: 0.0355
Epoch [5/6], Step [700/700], Loss: 0.0362
Epoch [6/6], Step [200/700], Loss: 0.0154
Epoch [6/6], Step [400/700], Loss: 0.0163
Epoch [6/6], Step [600/700], Loss: 0.0185
Epoch [6/6], Step [700/700], Loss: 0.0185
Test Accuracy of the model: 86.86 %
```

```
self.embedding=nn.Embedding(vocab_size, embedding_dim)
self.embedding.weight = nn.Parameter(torch.tensor(word_embedding,dtype=torch.float32))

self.lstm=nn.LSTM(embedding_dim, hidden_dim, n_layers, dropout=drop_prob, batch_first=True)

#dropout layer
self.dropout=nn.Dropout(0.3)

#Linear and sigmoid layer
self.fc1=nn.Linear(hidden_dim, 64)
self.fc2=nn.Linear(64, 16)
self.fc3=nn.Linear(16,output_size)
self.sigmoid=nn.Sigmoid()
```

Using reviews with length of the longest review and with 200 dimension in GloVe:

```
Epoch [1/6], Step [200/700], Loss: 0.6500
Epoch [1/6], Step [400/700], Loss: 0.6507
Epoch [1/6], Step [600/700], Loss: 0.5810
Epoch [1/6], Step [700/700], Loss: 0.5596
Epoch [2/6], Step [200/700], Loss: 0.3245
Epoch [2/6], Step [400/700], Loss: 0.2949
Epoch [2/6], Step [600/700], Loss: 0.2851
Epoch [2/6], Step [700/700], Loss: 0.2792
Epoch [3/6], Step [200/700], Loss: 0.1528
Epoch [3/6], Step [400/700], Loss: 0.1480
Epoch [3/6], Step [600/700], Loss: 0.1512
Epoch [3/6], Step [700/700], Loss: 0.1556
Epoch [4/6], Step [200/700], Loss: 0.0720
Epoch [4/6], Step [400/700], Loss: 0.0690
Epoch [4/6], Step [600/700], Loss: 0.0718
Epoch [4/6], Step [700/700], Loss: 0.0730
Epoch [5/6], Step [200/700], Loss: 0.0297
```

```
Epoch [5/6], Step [400/700], Loss: 0.0298
Epoch [5/6], Step [600/700], Loss: 0.0312
Epoch [5/6], Step [700/700], Loss: 0.0314
Epoch [6/6], Step [200/700], Loss: 0.0153
Epoch [6/6], Step [400/700], Loss: 0.0151
Epoch [6/6], Step [600/700], Loss: 0.0201
Epoch [6/6], Step [700/700], Loss: 0.0212
Test Accuracy of the model: 85.98666666666666 %
```

Using reviews with length of the longest review and with 300 dimension in GloVe:

```
Epoch [1/6], Step [200/700], Loss: 0.5820
Epoch [1/6], Step [400/700], Loss: 0.5045
Epoch [1/6], Step [600/700], Loss: 0.4651
Epoch [1/6], Step [700/700], Loss: 0.4489
Epoch [2/6], Step [200/700], Loss: 0.2512
Epoch [2/6], Step [400/700], Loss: 0.2341
Epoch [2/6], Step [600/700], Loss: 0.2360
Epoch [2/6], Step [700/700], Loss: 0.2371
Epoch [3/6], Step [200/700], Loss: 0.1079
Epoch [3/6], Step [400/700], Loss: 0.1179
Epoch [3/6], Step [600/700], Loss: 0.1194
Epoch [3/6], Step [700/700], Loss: 0.1188
Epoch [4/6], Step [200/700], Loss: 0.0486
Epoch [4/6], Step [400/700], Loss: 0.0477
Epoch [4/6], Step [600/700], Loss: 0.0489
Epoch [4/6], Step [700/700], Loss: 0.0510
Epoch [5/6], Step [200/700], Loss: 0.0159
Epoch [5/6], Step [400/700], Loss: 0.0209
Epoch [5/6], Step [600/700], Loss: 0.0219
Epoch [5/6], Step [700/700], Loss: 0.0231
Epoch [6/6], Step [200/700], Loss: 0.0086
Epoch [6/6], Step [400/700], Loss: 0.0111
Epoch [6/6], Step [600/700], Loss: 0.0127
Epoch [6/6], Step [700/700], Loss: 0.0133
Test Accuracy of the model: 86.58 %
```

We can see that the file with 100 dimensions in GloVe gives te best accuracy. So we are going to proceed with it.

Using reviews with length of the longest review and 128 hidden dimension:

```
Epoch [1/6], Step [200/700], Loss: 0.6490
Epoch [1/6], Step [400/700], Loss: 0.5561
Epoch [1/6], Step [600/700], Loss: 0.5042
```

```
Epoch [1/6], Step [700/700], Loss: 0.4864
Epoch [2/6], Step [200/700], Loss: 0.2751
Epoch [2/6], Step [400/700], Loss: 0.2655
Epoch [2/6], Step [600/700], Loss: 0.2557
Epoch [2/6], Step [700/700], Loss: 0.2523
Epoch [3/6], Step [200/700], Loss: 0.1238
Epoch [3/6], Step [400/700], Loss: 0.1255
Epoch [3/6], Step [600/700], Loss: 0.1286
Epoch [3/6], Step [700/700], Loss: 0.1308
Epoch [4/6], Step [200/700], Loss: 0.0528
Epoch [4/6], Step [400/700], Loss: 0.0564
Epoch [4/6], Step [600/700], Loss: 0.0581
Epoch [4/6], Step [700/700], Loss: 0.0581
Epoch [5/6], Step [200/700], Loss: 0.0216
Epoch [5/6], Step [400/700], Loss: 0.0241
Epoch [5/6], Step [600/700], Loss: 0.0225
Epoch [5/6], Step [700/700], Loss: 0.0224
Epoch [6/6], Step [200/700], Loss: 0.0103
Epoch [6/6], Step [400/700], Loss: 0.0169
Epoch [6/6], Step [600/700], Loss: 0.0183
Epoch [6/6], Step [700/700], Loss: 0.0176
Test Accuracy of the model: 86.72 %
```

Using reviews with length of the longest review and 1 layer lstm:

```
Epoch [1/6], Step [200/700], Loss: 0.5974
Epoch [1/6], Step [400/700], Loss: 0.5294
Epoch [1/6], Step [600/700], Loss: 0.4930
Epoch [1/6], Step [700/700], Loss: 0.4741
Epoch [2/6], Step [200/700], Loss: 0.2909
Epoch [2/6], Step [400/700], Loss: 0.2780
Epoch [2/6], Step [600/700], Loss: 0.2602
Epoch [2/6], Step [700/700], Loss: 0.2585
Epoch [3/6], Step [200/700], Loss: 0.1261
Epoch [3/6], Step [400/700], Loss: 0.1269
Epoch [3/6], Step [600/700], Loss: 0.1256
Epoch [3/6], Step [700/700], Loss: 0.1257
Epoch [4/6], Step [200/700], Loss: 0.0511
Epoch [4/6], Step [400/700], Loss: 0.0507
Epoch [4/6], Step [600/700], Loss: 0.0543
Epoch [4/6], Step [700/700], Loss: 0.0541
Epoch [5/6], Step [200/700], Loss: 0.0441
Epoch [5/6], Step [400/700], Loss: 0.0337
Epoch [5/6], Step [600/700], Loss: 0.0309
```

```
Epoch [5/6], Step [700/700], Loss: 0.0308
Epoch [6/6], Step [200/700], Loss: 0.0164
Epoch [6/6], Step [400/700], Loss: 0.0140
Epoch [6/6], Step [600/700], Loss: 0.0182
Epoch [6/6], Step [700/700], Loss: 0.0169
Test Accuracy of the model: 86.12666666666667 %
```

With 3 lstm layers or more we get worse results.

Using reviews with length of the longest review changing the architecture:

```
Epoch [1/6], Step [200/700], Loss: 0.6007
Epoch [1/6], Step [400/700], Loss: 0.5384
Epoch [1/6], Step [600/700], Loss: 0.5033
Epoch [1/6], Step [700/700], Loss: 0.4865
Epoch [2/6], Step [200/700], Loss: 0.2834
Epoch [2/6], Step [400/700], Loss: 0.2756
Epoch [2/6], Step [600/700], Loss: 0.2653
Epoch [2/6], Step [700/700], Loss: 0.2656
Epoch [3/6], Step [200/700], Loss: 0.1447
Epoch [3/6], Step [400/700], Loss: 0.1361
Epoch [3/6], Step [600/700], Loss: 0.1351
Epoch [3/6], Step [700/700], Loss: 0.1364
Epoch [4/6], Step [200/700], Loss: 0.0589
Epoch [4/6], Step [400/700], Loss: 0.0610
Epoch [4/6], Step [600/700], Loss: 0.0657
Epoch [4/6], Step [700/700], Loss: 0.0659
Epoch [5/6], Step [200/700], Loss: 0.0374
Epoch [5/6], Step [400/700], Loss: 0.0385
Epoch [5/6], Step [600/700], Loss: 0.0361
Epoch [5/6], Step [700/700], Loss: 0.0356
Epoch [6/6], Step [200/700], Loss: 0.0163
Epoch [6/6], Step [400/700], Loss: 0.0193
Epoch [6/6], Step [600/700], Loss: 0.0258
Epoch [6/6], Step [700/700], Loss: 0.0278
Test Accuracy of the model: 86.81333333333333 %
```

```
self.embedding=nn.Embedding(vocab_size, embedding_dim)
self.embedding.weight = nn.Parameter(torch.tensor(word_embedding,dtype=torch.float32))

self.lstm=nn.LSTM(embedding_dim, hidden_dim, n_layers, dropout=drop_prob, batch_first=True)

#dropout layer
self.dropout=nn.Dropout(0.5)

#Linear and sigmoid layer
self.fc1=nn.Linear(hidden_dim, 128)
self.fc2=nn.Linear(128, 64)
self.fc3=nn.Linear(64, 32)
self.fc4=nn.Linear(32, 16)
self.fc5=nn.Linear(16,output_size)
self.sigmoid=nn.Sigmoid()
```

We decide to keep padding with the architecture of the beginning, number of lstm layers = 2 and hidden dimension 256. After that we start to try different values to the hyperparameters of the optimizers: learning rate for Adam and learning rate and momentum for SGD.

Padding with Adam using GloVe embedding:

1. lr = 0,001

```
Epoch [1/6], Step [200/700], Loss: 0.5771
Epoch [1/6], Step [400/700], Loss: 0.5288
Epoch [1/6], Step [600/700], Loss: 0.4875
Epoch [1/6], Step [700/700], Loss: 0.4830
Epoch [2/6], Step [200/700], Loss: 0.3211
Epoch [2/6], Step [400/700], Loss: 0.3021
Epoch [2/6], Step [600/700], Loss: 0.2974
Epoch [2/6], Step [700/700], Loss: 0.2888
Epoch [3/6], Step [200/700], Loss: 0.1556
Epoch [3/6], Step [400/700], Loss: 0.1580
Epoch [3/6], Step [600/700], Loss: 0.1596
Epoch [3/6], Step [700/700], Loss: 0.1602
Epoch [4/6], Step [200/700], Loss: 0.0803
Epoch [4/6], Step [400/700], Loss: 0.0751
Epoch [4/6], Step [600/700], Loss: 0.0799
Epoch [4/6], Step [700/700], Loss: 0.0804
Epoch [5/6], Step [200/700], Loss: 0.0395
Epoch [5/6], Step [400/700], Loss: 0.0389
Epoch [5/6], Step [600/700], Loss: 0.0396
Epoch [5/6], Step [700/700], Loss: 0.0459
Epoch [6/6], Step [200/700], Loss: 0.0698
Epoch [6/6], Step [400/700], Loss: 0.0560
Epoch [6/6], Step [600/700], Loss: 0.0516
Epoch [6/6], Step [700/700], Loss: 0.0499
Test Accuracy of the model: 86.23333333333333 %
```

2. lr = 0,0005

```
Epoch [1/6], Step [200/700], Loss: 0.6501
Epoch [1/6], Step [400/700], Loss: 0.5942
Epoch [1/6], Step [600/700], Loss: 0.5522
Epoch [1/6], Step [700/700], Loss: 0.5323
Epoch [2/6], Step [200/700], Loss: 0.3411
Epoch [2/6], Step [400/700], Loss: 0.3316
Epoch [2/6], Step [600/700], Loss: 0.3259
Epoch [2/6], Step [700/700], Loss: 0.3243
Epoch [3/6], Step [200/700], Loss: 0.2506
Epoch [3/6], Step [400/700], Loss: 0.2462
Epoch [3/6], Step [600/700], Loss: 0.2418
Epoch [3/6], Step [700/700], Loss: 0.2416
Epoch [4/6], Step [200/700], Loss: 0.1649
Epoch [4/6], Step [400/700], Loss: 0.1655
Epoch [4/6], Step [600/700], Loss: 0.1666
Epoch [4/6], Step [700/700], Loss: 0.1675
Epoch [5/6], Step [200/700], Loss: 0.1147
Epoch [5/6], Step [400/700], Loss: 0.1154
Epoch [5/6], Step [600/700], Loss: 0.1157
Epoch [5/6], Step [700/700], Loss: 0.1154
Epoch [6/6], Step [200/700], Loss: 0.0686
Epoch [6/6], Step [400/700], Loss: 0.0700
Epoch [6/6], Step [600/700], Loss: 0.0691
Epoch [6/6], Step [700/700], Loss: 0.0683
Test Accuracy of the model: 87.5 %
```

3. lr = 0,002

```
Epoch [1/6], Step [200/700], Loss: 0.6829
Epoch [1/6], Step [400/700], Loss: 0.6515
Epoch [1/6], Step [600/700], Loss: 0.5605
Epoch [1/6], Step [700/700], Loss: 0.5301
Epoch [2/6], Step [200/700], Loss: 0.2533
Epoch [2/6], Step [400/700], Loss: 0.2444
Epoch [2/6], Step [600/700], Loss: 0.2390
Epoch [2/6], Step [700/700], Loss: 0.2379
Epoch [3/6], Step [200/700], Loss: 0.1073
Epoch [3/6], Step [400/700], Loss: 0.1070
Epoch [3/6], Step [600/700], Loss: 0.1074
Epoch [3/6], Step [700/700], Loss: 0.1098
Epoch [4/6], Step [200/700], Loss: 0.0440
Epoch [4/6], Step [400/700], Loss: 0.0474
Epoch [4/6], Step [600/700], Loss: 0.0506
```

```
Epoch [4/6], Step [700/700], Loss: 0.0500
Epoch [5/6], Step [200/700], Loss: 0.0319
Epoch [5/6], Step [400/700], Loss: 0.0235
Epoch [5/6], Step [600/700], Loss: 0.0247
Epoch [5/6], Step [700/700], Loss: 0.0255
Epoch [6/6], Step [200/700], Loss: 0.0211
Epoch [6/6], Step [400/700], Loss: 0.0215
Epoch [6/6], Step [600/700], Loss: 0.0188
Epoch [6/6], Step [700/700], Loss: 0.0211
Test Accuracy of the model: 87.52666666666667 %
```

Padding with SGD:

1. lr = 0,001 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6942
Epoch [1/6], Step [400/700], Loss: 0.6940
Epoch [1/6], Step [600/700], Loss: 0.6939
Epoch [1/6], Step [700/700], Loss: 0.6938
Epoch [2/6], Step [200/700], Loss: 0.6934
Epoch [2/6], Step [400/700], Loss: 0.6932
Epoch [2/6], Step [600/700], Loss: 0.6931
Epoch [2/6], Step [700/700], Loss: 0.6931
Epoch [3/6], Step [200/700], Loss: 0.6935
Epoch [3/6], Step [400/700], Loss: 0.6932
Epoch [3/6], Step [600/700], Loss: 0.6931
Epoch [3/6], Step [700/700], Loss: 0.6930
Epoch [4/6], Step [200/700], Loss: 0.6926
Epoch [4/6], Step [400/700], Loss: 0.6928
Epoch [4/6], Step [600/700], Loss: 0.6928
Epoch [4/6], Step [700/700], Loss: 0.6928
Epoch [5/6], Step [200/700], Loss: 0.6928
Epoch [5/6], Step [400/700], Loss: 0.6928
Epoch [5/6], Step [600/700], Loss: 0.6927
Epoch [5/6], Step [700/700], Loss: 0.6928
Epoch [6/6], Step [200/700], Loss: 0.6923
Epoch [6/6], Step [400/700], Loss: 0.6923
Epoch [6/6], Step [600/700], Loss: 0.6924
Epoch [6/6], Step [700/700], Loss: 0.6923
Test Accuracy of the model: 56.38666666666666 %
```

2. lr = 0,01 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6941
Epoch [1/6], Step [400/700], Loss: 0.6936
Epoch [1/6], Step [600/700], Loss: 0.6934
```

```
Epoch [1/6], Step [700/700], Loss: 0.6933
Epoch [2/6], Step [200/700], Loss: 0.6922
Epoch [2/6], Step [400/700], Loss: 0.6915
Epoch [2/6], Step [600/700], Loss: 0.6900
Epoch [2/6], Step [700/700], Loss: 0.6881
Epoch [3/6], Step [200/700], Loss: 0.6242
Epoch [3/6], Step [400/700], Loss: 0.6206
Epoch [3/6], Step [600/700], Loss: 0.6193
Epoch [3/6], Step [700/700], Loss: 0.6134
Epoch [4/6], Step [200/700], Loss: 0.6420
Epoch [4/6], Step [400/700], Loss: 0.6296
Epoch [4/6], Step [600/700], Loss: 0.6083
Epoch [4/6], Step [700/700], Loss: 0.6002
Epoch [5/6], Step [200/700], Loss: 0.5540
Epoch [5/6], Step [400/700], Loss: 0.5367
Epoch [5/6], Step [600/700], Loss: 0.5235
Epoch [5/6], Step [700/700], Loss: 0.5206
Epoch [6/6], Step [200/700], Loss: 0.4658
Epoch [6/6], Step [400/700], Loss: 0.4518
Epoch [6/6], Step [600/700], Loss: 0.4384
Epoch [6/6], Step [700/700], Loss: 0.4332
Test Accuracy of the model: 82.48666666666666 %
```

3. lr = 0,05 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6939
Epoch [1/6], Step [400/700], Loss: 0.6848
Epoch [1/6], Step [600/700], Loss: 0.6637
Epoch [1/6], Step [700/700], Loss: 0.6545
Epoch [2/6], Step [200/700], Loss: 0.5673
Epoch [2/6], Step [400/700], Loss: 0.5836
Epoch [2/6], Step [600/700], Loss: 0.5607
Epoch [2/6], Step [700/700], Loss: 0.5496
Epoch [3/6], Step [200/700], Loss: 0.6860
Epoch [3/6], Step [400/700], Loss: 0.6265
Epoch [3/6], Step [600/700], Loss: 0.5721
Epoch [3/6], Step [700/700], Loss: 0.5485
Epoch [4/6], Step [200/700], Loss: 0.3917
Epoch [4/6], Step [400/700], Loss: 0.3778
Epoch [4/6], Step [600/700], Loss: 0.3709
Epoch [4/6], Step [700/700], Loss: 0.3701
Epoch [5/6], Step [200/700], Loss: 0.3442
Epoch [5/6], Step [400/700], Loss: 0.3383
Epoch [5/6], Step [600/700], Loss: 0.3374
```

```
Epoch [5/6], Step [700/700], Loss: 0.3354
Epoch [6/6], Step [200/700], Loss: 0.3181
Epoch [6/6], Step [400/700], Loss: 0.3184
Epoch [6/6], Step [600/700], Loss: 0.3175
Epoch [6/6], Step [700/700], Loss: 0.3150
Test Accuracy of the model: 87.81333333333333 %
```

4. lr = 0,06 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6930
Epoch [1/6], Step [400/700], Loss: 0.6816
Epoch [1/6], Step [600/700], Loss: 0.6630
Epoch [1/6], Step [700/700], Loss: 0.6538
Epoch [2/6], Step [200/700], Loss: 0.6503
Epoch [2/6], Step [400/700], Loss: 0.6358
Epoch [2/6], Step [600/700], Loss: 0.6286
Epoch [2/6], Step [700/700], Loss: 0.6158
Epoch [3/6], Step [200/700], Loss: 0.5656
Epoch [3/6], Step [400/700], Loss: 0.5026
Epoch [3/6], Step [600/700], Loss: 0.4666
Epoch [3/6], Step [700/700], Loss: 0.4526
Epoch [4/6], Step [200/700], Loss: 0.3702
Epoch [4/6], Step [400/700], Loss: 0.3646
Epoch [4/6], Step [600/700], Loss: 0.3591
Epoch [4/6], Step [700/700], Loss: 0.3542
Epoch [5/6], Step [200/700], Loss: 0.3218
Epoch [5/6], Step [400/700], Loss: 0.3326
Epoch [5/6], Step [600/700], Loss: 0.3278
Epoch [5/6], Step [700/700], Loss: 0.3276
Epoch [6/6], Step [200/700], Loss: 0.3250
Epoch [6/6], Step [400/700], Loss: 0.3090
Epoch [6/6], Step [600/700], Loss: 0.3076
Epoch [6/6], Step [700/700], Loss: 0.3045
Test Accuracy of the model: 87.93333333333334 %
```

5. lr = 0,06 and mom = 0,8

```
Epoch [1/6], Step [200/700], Loss: 0.6945
Epoch [1/6], Step [400/700], Loss: 0.6934
Epoch [1/6], Step [600/700], Loss: 0.6846
Epoch [1/6], Step [700/700], Loss: 0.6762
Epoch [2/6], Step [200/700], Loss: 0.5848
Epoch [2/6], Step [400/700], Loss: 0.5760
Epoch [2/6], Step [600/700], Loss: 0.5789
Epoch [2/6], Step [700/700], Loss: 0.5734
```

```
Epoch [3/6], Step [200/700], Loss: 0.5180
Epoch [3/6], Step [400/700], Loss: 0.4927
Epoch [3/6], Step [600/700], Loss: 0.4737
Epoch [3/6], Step [700/700], Loss: 0.4642
Epoch [4/6], Step [200/700], Loss: 0.3850
Epoch [4/6], Step [400/700], Loss: 0.3758
Epoch [4/6], Step [600/700], Loss: 0.3692
Epoch [4/6], Step [700/700], Loss: 0.3663
Epoch [5/6], Step [200/700], Loss: 0.3398
Epoch [5/6], Step [400/700], Loss: 0.3356
Epoch [5/6], Step [600/700], Loss: 0.3374
Epoch [5/6], Step [700/700], Loss: 0.3340
Epoch [6/6], Step [200/700], Loss: 0.3219
Epoch [6/6], Step [400/700], Loss: 0.3169
Epoch [6/6], Step [600/700], Loss: 0.3103
Epoch [6/6], Step [700/700], Loss: 0.3103
Test Accuracy of the model: 87.52 %
```

6. lr = 0,06 and mom = 0,85

```
Epoch [1/6], Step [200/700], Loss: 0.6944
Epoch [1/6], Step [400/700], Loss: 0.6941
Epoch [1/6], Step [600/700], Loss: 0.6931
Epoch [1/6], Step [700/700], Loss: 0.6869
Epoch [2/6], Step [200/700], Loss: 0.6103
Epoch [2/6], Step [400/700], Loss: 0.6047
Epoch [2/6], Step [600/700], Loss: 0.5803
Epoch [2/6], Step [700/700], Loss: 0.5647
Epoch [3/6], Step [200/700], Loss: 0.4086
Epoch [3/6], Step [400/700], Loss: 0.3999
Epoch [3/6], Step [600/700], Loss: 0.3890
Epoch [3/6], Step [700/700], Loss: 0.3857
Epoch [4/6], Step [200/700], Loss: 0.3479
Epoch [4/6], Step [400/700], Loss: 0.3471
Epoch [4/6], Step [600/700], Loss: 0.3442
Epoch [4/6], Step [700/700], Loss: 0.3419
Epoch [5/6], Step [200/700], Loss: 0.3156
Epoch [5/6], Step [400/700], Loss: 0.3211
Epoch [5/6], Step [600/700], Loss: 0.3213
Epoch [5/6], Step [700/700], Loss: 0.3210
Epoch [6/6], Step [200/700], Loss: 0.3024
Epoch [6/6], Step [400/700], Loss: 0.3027
Epoch [6/6], Step [600/700], Loss: 0.3025
Epoch [6/6], Step [700/700], Loss: 0.3048
```

```
Test Accuracy of the model: 87.02666666666667 %
```

7. lr = 0,06 and mom = 0,89

```
Epoch [1/6], Step [200/700], Loss: 0.6942
Epoch [1/6], Step [400/700], Loss: 0.6885
Epoch [1/6], Step [600/700], Loss: 0.6649
Epoch [1/6], Step [700/700], Loss: 0.6561
Epoch [2/6], Step [200/700], Loss: 0.6558
Epoch [2/6], Step [400/700], Loss: 0.6222
Epoch [2/6], Step [600/700], Loss: 0.5700
Epoch [2/6], Step [700/700], Loss: 0.5459
Epoch [3/6], Step [200/700], Loss: 0.3867
Epoch [3/6], Step [400/700], Loss: 0.3803
Epoch [3/6], Step [600/700], Loss: 0.3734
Epoch [3/6], Step [700/700], Loss: 0.3722
Epoch [4/6], Step [200/700], Loss: 0.3386
Epoch [4/6], Step [400/700], Loss: 0.3383
Epoch [4/6], Step [600/700], Loss: 0.3351
Epoch [4/6], Step [700/700], Loss: 0.3359
Epoch [5/6], Step [200/700], Loss: 0.3180
Epoch [5/6], Step [400/700], Loss: 0.3170
Epoch [5/6], Step [600/700], Loss: 0.3144
Epoch [5/6], Step [700/700], Loss: 0.3149
Epoch [6/6], Step [200/700], Loss: 0.3026
Epoch [6/6], Step [400/700], Loss: 0.3003
Epoch [6/6], Step [600/700], Loss: 0.2981
Epoch [6/6], Step [700/700], Loss: 0.3017
Test Accuracy of the model: 86.62666666666667 %
```

8. lr = 0,06 and mom = 0,95

```
Epoch [1/6], Step [200/700], Loss: 0.6947
Epoch [1/6], Step [400/700], Loss: 0.6675
Epoch [1/6], Step [600/700], Loss: 0.6531
Epoch [1/6], Step [700/700], Loss: 0.6492
Epoch [2/6], Step [200/700], Loss: 0.4959
Epoch [2/6], Step [400/700], Loss: 0.4639
Epoch [2/6], Step [600/700], Loss: 0.4568
Epoch [2/6], Step [700/700], Loss: 0.4469
Epoch [3/6], Step [200/700], Loss: 0.3806
Epoch [3/6], Step [400/700], Loss: 0.3669
Epoch [3/6], Step [600/700], Loss: 0.3647
Epoch [3/6], Step [700/700], Loss: 0.3624
Epoch [4/6], Step [200/700], Loss: 0.3266
```

```
Epoch [4/6], Step [400/700], Loss: 0.3230
Epoch [4/6], Step [600/700], Loss: 0.3224
Epoch [4/6], Step [700/700], Loss: 0.3240
Epoch [5/6], Step [200/700], Loss: 0.2820
Epoch [5/6], Step [400/700], Loss: 0.2985
Epoch [5/6], Step [600/700], Loss: 0.2941
Epoch [5/6], Step [700/700], Loss: 0.2922
Epoch [6/6], Step [200/700], Loss: 0.2860
Epoch [6/6], Step [400/700], Loss: 0.2888
Epoch [6/6], Step [600/700], Loss: 0.2860
Epoch [6/6], Step [700/700], Loss: 0.2860
Test Accuracy of the model: 89.02666666666667 %
```

Padding with Adam using no pretrained embedding:

1. lr = 0,001

```
Epoch [1/6], Step [200/700], Loss: 0.6488
Epoch [1/6], Step [400/700], Loss: 0.6151
Epoch [1/6], Step [600/700], Loss: 0.6418
Epoch [1/6], Step [700/700], Loss: 0.6492
Epoch [2/6], Step [200/700], Loss: 0.6860
Epoch [2/6], Step [400/700], Loss: 0.6406
Epoch [2/6], Step [600/700], Loss: 0.5756
Epoch [2/6], Step [700/700], Loss: 0.5499
Epoch [3/6], Step [200/700], Loss: 0.3475
Epoch [3/6], Step [400/700], Loss: 0.3327
Epoch [3/6], Step [600/700], Loss: 0.3270
Epoch [3/6], Step [700/700], Loss: 0.3232
Epoch [4/6], Step [200/700], Loss: 0.2342
Epoch [4/6], Step [400/700], Loss: 0.2345
Epoch [4/6], Step [600/700], Loss: 0.2317
Epoch [4/6], Step [700/700], Loss: 0.2308
Epoch [5/6], Step [200/700], Loss: 0.1700
Epoch [5/6], Step [400/700], Loss: 0.1728
Epoch [5/6], Step [600/700], Loss: 0.1725
Epoch [5/6], Step [700/700], Loss: 0.1739
Epoch [6/6], Step [200/700], Loss: 0.1287
Epoch [6/6], Step [400/700], Loss: 0.1236
Epoch [6/6], Step [600/700], Loss: 0.1272
Epoch [6/6], Step [700/700], Loss: 0.1285
Test Accuracy of the model: 87.94 %
```

2. lr = 0,0005

```
Epoch [1/6], Step [200/700], Loss: 0.6631
```

```
Epoch [1/6], Step [400/700], Loss: 0.6108
Epoch [1/6], Step [600/700], Loss: 0.5737
Epoch [1/6], Step [700/700], Loss: 0.5664
Epoch [2/6], Step [200/700], Loss: 0.4614
Epoch [2/6], Step [400/700], Loss: 0.4301
Epoch [2/6], Step [600/700], Loss: 0.4163
Epoch [2/6], Step [700/700], Loss: 0.4159
Epoch [3/6], Step [200/700], Loss: 0.3787
Epoch [3/6], Step [400/700], Loss: 0.3622
Epoch [3/6], Step [600/700], Loss: 0.3552
Epoch [3/6], Step [700/700], Loss: 0.3545
Epoch [4/6], Step [200/700], Loss: 0.3061
Epoch [4/6], Step [400/700], Loss: 0.2972
Epoch [4/6], Step [600/700], Loss: 0.3027
Epoch [4/6], Step [700/700], Loss: 0.3016
Epoch [5/6], Step [200/700], Loss: 0.2464
Epoch [5/6], Step [400/700], Loss: 0.2924
Epoch [5/6], Step [600/700], Loss: 0.2977
Epoch [5/6], Step [700/700], Loss: 0.2949
Epoch [6/6], Step [200/700], Loss: 0.2446
Epoch [6/6], Step [400/700], Loss: 0.2502
Epoch [6/6], Step [600/700], Loss: 0.2474
Epoch [6/6], Step [700/700], Loss: 0.2598
Test Accuracy of the model: 85.02 %
```

3. lr = 0,002

```
Epoch [1/6], Step [200/700], Loss: 0.6410
Epoch [1/6], Step [400/700], Loss: 0.6493
Epoch [1/6], Step [600/700], Loss: 0.6644
Epoch [1/6], Step [700/700], Loss: 0.6684
Epoch [2/6], Step [200/700], Loss: 0.6881
Epoch [2/6], Step [400/700], Loss: 0.6406
Epoch [2/6], Step [600/700], Loss: 0.5844
Epoch [2/6], Step [700/700], Loss: 0.5571
Epoch [3/6], Step [200/700], Loss: 0.3215
Epoch [3/6], Step [400/700], Loss: 0.3157
Epoch [3/6], Step [600/700], Loss: 0.3107
Epoch [3/6], Step [700/700], Loss: 0.3066
Epoch [4/6], Step [200/700], Loss: 0.1947
Epoch [4/6], Step [400/700], Loss: 0.2008
Epoch [4/6], Step [600/700], Loss: 0.1977
Epoch [4/6], Step [700/700], Loss: 0.1982
Epoch [5/6], Step [200/700], Loss: 0.1292
```

```
Epoch [5/6], Step [400/700], Loss: 0.1302
Epoch [5/6], Step [600/700], Loss: 0.1336
Epoch [5/6], Step [700/700], Loss: 0.1338
Epoch [6/6], Step [200/700], Loss: 0.0804
Epoch [6/6], Step [400/700], Loss: 0.0825
Epoch [6/6], Step [600/700], Loss: 0.0877
Epoch [6/6], Step [700/700], Loss: 0.0890
Test Accuracy of the model: 87.26 %
```

Padding with SGD:

1. lr = 0,001 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6952
Epoch [1/6], Step [400/700], Loss: 0.6945
Epoch [1/6], Step [600/700], Loss: 0.6943
Epoch [1/6], Step [700/700], Loss: 0.6942
Epoch [2/6], Step [200/700], Loss: 0.6932
Epoch [2/6], Step [400/700], Loss: 0.6933
Epoch [2/6], Step [600/700], Loss: 0.6934
Epoch [2/6], Step [700/700], Loss: 0.6934
Epoch [3/6], Step [200/700], Loss: 0.6933
Epoch [3/6], Step [400/700], Loss: 0.6934
Epoch [3/6], Step [600/700], Loss: 0.6933
Epoch [3/6], Step [700/700], Loss: 0.6932
Epoch [4/6], Step [200/700], Loss: 0.6934
Epoch [4/6], Step [400/700], Loss: 0.6933
Epoch [4/6], Step [600/700], Loss: 0.6933
Epoch [4/6], Step [700/700], Loss: 0.6934
Epoch [5/6], Step [200/700], Loss: 0.6931
Epoch [5/6], Step [400/700], Loss: 0.6932
Epoch [5/6], Step [600/700], Loss: 0.6933
Epoch [5/6], Step [700/700], Loss: 0.6933
Epoch [6/6], Step [200/700], Loss: 0.6931
Epoch [6/6], Step [400/700], Loss: 0.6932
Epoch [6/6], Step [600/700], Loss: 0.6931
Epoch [6/6], Step [700/700], Loss: 0.6931
Test Accuracy of the model: 49.58 %
```

2. lr = 0,01 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6938
Epoch [1/6], Step [400/700], Loss: 0.6936
Epoch [1/6], Step [600/700], Loss: 0.6936
Epoch [1/6], Step [700/700], Loss: 0.6935
Epoch [2/6], Step [200/700], Loss: 0.6933
```

```
Epoch [2/6], Step [400/700], Loss: 0.6930
Epoch [2/6], Step [600/700], Loss: 0.6928
Epoch [2/6], Step [700/700], Loss: 0.6925
Epoch [3/6], Step [200/700], Loss: 0.6909
Epoch [3/6], Step [400/700], Loss: 0.6895
Epoch [3/6], Step [600/700], Loss: 0.6875
Epoch [3/6], Step [700/700], Loss: 0.6860
Epoch [4/6], Step [200/700], Loss: 0.6673
Epoch [4/6], Step [400/700], Loss: 0.6640
Epoch [4/6], Step [600/700], Loss: 0.6605
Epoch [4/6], Step [700/700], Loss: 0.6566
Epoch [5/6], Step [200/700], Loss: 0.6170
Epoch [5/6], Step [400/700], Loss: 0.6130
Epoch [5/6], Step [600/700], Loss: 0.6002
Epoch [5/6], Step [700/700], Loss: 0.5955
Epoch [6/6], Step [200/700], Loss: 0.5656
Epoch [6/6], Step [400/700], Loss: 0.5424
Epoch [6/6], Step [600/700], Loss: 0.5355
Epoch [6/6], Step [700/700], Loss: 0.5288
Test Accuracy of the model: 75.27333333333333 %
```

3. lr = 0,05 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6938
Epoch [1/6], Step [400/700], Loss: 0.6915
Epoch [1/6], Step [600/700], Loss: 0.6820
Epoch [1/6], Step [700/700], Loss: 0.6753
Epoch [2/6], Step [200/700], Loss: 0.6315
Epoch [2/6], Step [400/700], Loss: 0.6450
Epoch [2/6], Step [600/700], Loss: 0.6472
Epoch [2/6], Step [700/700], Loss: 0.6431
Epoch [3/6], Step [200/700], Loss: 0.6979
Epoch [3/6], Step [400/700], Loss: 0.6851
Epoch [3/6], Step [600/700], Loss: 0.6636
Epoch [3/6], Step [700/700], Loss: 0.6546
Epoch [4/6], Step [200/700], Loss: 0.6461
Epoch [4/6], Step [400/700], Loss: 0.6123
Epoch [4/6], Step [600/700], Loss: 0.5831
Epoch [4/6], Step [700/700], Loss: 0.5686
Epoch [5/6], Step [200/700], Loss: 0.4527
Epoch [5/6], Step [400/700], Loss: 0.4456
Epoch [5/6], Step [600/700], Loss: 0.4376
Epoch [5/6], Step [700/700], Loss: 0.4343
Epoch [6/6], Step [200/700], Loss: 0.3983
```

```
Epoch [6/6], Step [400/700], Loss: 0.3927
Epoch [6/6], Step [600/700], Loss: 0.3852
Epoch [6/6], Step [700/700], Loss: 0.3835
Test Accuracy of the model: 83.18666666666667 %
```

4. lr = 0,06 and mom = 0,9

```
Epoch [1/6], Step [200/700], Loss: 0.6945
Epoch [1/6], Step [400/700], Loss: 0.6935
Epoch [1/6], Step [600/700], Loss: 0.6848
Epoch [1/6], Step [700/700], Loss: 0.6794
Epoch [2/6], Step [200/700], Loss: 0.6388
Epoch [2/6], Step [400/700], Loss: 0.6319
Epoch [2/6], Step [600/700], Loss: 0.6191
Epoch [2/6], Step [700/700], Loss: 0.6202
Epoch [3/6], Step [200/700], Loss: 0.5597
Epoch [3/6], Step [400/700], Loss: 0.5711
Epoch [3/6], Step [600/700], Loss: 0.5882
Epoch [3/6], Step [700/700], Loss: 0.5844
Epoch [4/6], Step [200/700], Loss: 0.5247
Epoch [4/6], Step [400/700], Loss: 0.5315
Epoch [4/6], Step [600/700], Loss: 0.5141
Epoch [4/6], Step [700/700], Loss: 0.5108
Epoch [5/6], Step [200/700], Loss: 0.4399
Epoch [5/6], Step [400/700], Loss: 0.4333
Epoch [5/6], Step [600/700], Loss: 0.4235
Epoch [5/6], Step [700/700], Loss: 0.4185
Epoch [6/6], Step [200/700], Loss: 0.3801
Epoch [6/6], Step [400/700], Loss: 0.3775
Epoch [6/6], Step [600/700], Loss: 0.3752
Epoch [6/6], Step [700/700], Loss: 0.3821
Test Accuracy of the model: 82.63333333333334 %
```

5. lr = 0,05 and mom = 0,8

```
Epoch [1/6], Step [200/700], Loss: 0.6941
Epoch [1/6], Step [400/700], Loss: 0.6932
Epoch [1/6], Step [600/700], Loss: 0.6929
Epoch [1/6], Step [700/700], Loss: 0.6919
Epoch [2/6], Step [200/700], Loss: 0.6705
Epoch [2/6], Step [400/700], Loss: 0.6590
Epoch [2/6], Step [600/700], Loss: 0.6531
Epoch [2/6], Step [700/700], Loss: 0.6523
Epoch [3/6], Step [200/700], Loss: 0.6161
Epoch [3/6], Step [400/700], Loss: 0.6089
```

```
Epoch [3/6], Step [600/700], Loss: 0.6089
Epoch [3/6], Step [700/700], Loss: 0.6059
Epoch [4/6], Step [200/700], Loss: 0.5675
Epoch [4/6], Step [400/700], Loss: 0.5702
Epoch [4/6], Step [600/700], Loss: 0.5763
Epoch [4/6], Step [700/700], Loss: 0.5769
Epoch [5/6], Step [200/700], Loss: 0.6953
Epoch [5/6], Step [400/700], Loss: 0.6944
Epoch [5/6], Step [600/700], Loss: 0.6941
Epoch [5/6], Step [700/700], Loss: 0.6940
Epoch [6/6], Step [200/700], Loss: 0.6937
Epoch [6/6], Step [400/700], Loss: 0.6933
Epoch [6/6], Step [600/700], Loss: 0.6927
Epoch [6/6], Step [700/700], Loss: 0.6918
Test Accuracy of the model: 58.153333333333336 %
```

6. lr = 0,05 and mom = 0,85

```
Epoch [1/6], Step [200/700], Loss: 0.6937
Epoch [1/6], Step [400/700], Loss: 0.6934
Epoch [1/6], Step [600/700], Loss: 0.6923
Epoch [1/6], Step [700/700], Loss: 0.6901
Epoch [2/6], Step [200/700], Loss: 0.6575
Epoch [2/6], Step [400/700], Loss: 0.6469
Epoch [2/6], Step [600/700], Loss: 0.6371
Epoch [2/6], Step [700/700], Loss: 0.6303
Epoch [3/6], Step [200/700], Loss: 0.5630
Epoch [3/6], Step [400/700], Loss: 0.5450
Epoch [3/6], Step [600/700], Loss: 0.5409
Epoch [3/6], Step [700/700], Loss: 0.5388
Epoch [4/6], Step [200/700], Loss: 0.5060
Epoch [4/6], Step [400/700], Loss: 0.4846
Epoch [4/6], Step [600/700], Loss: 0.4827
Epoch [4/6], Step [700/700], Loss: 0.4807
Epoch [5/6], Step [200/700], Loss: 0.4450
Epoch [5/6], Step [400/700], Loss: 0.4611
Epoch [5/6], Step [600/700], Loss: 0.4518
Epoch [5/6], Step [700/700], Loss: 0.4497
Epoch [6/6], Step [200/700], Loss: 0.4325
Epoch [6/6], Step [400/700], Loss: 0.4361
Epoch [6/6], Step [600/700], Loss: 0.5038
Epoch [6/6], Step [700/700], Loss: 0.5285
Test Accuracy of the model: 62.39333333333333 %
```

7. lr = 0,05 and mom = 0,89

```
Epoch [1/6], Step [200/700], Loss: 0.6942
Epoch [1/6], Step [400/700], Loss: 0.6936
Epoch [1/6], Step [600/700], Loss: 0.6909
Epoch [1/6], Step [700/700], Loss: 0.6876
Epoch [2/6], Step [200/700], Loss: 0.6511
Epoch [2/6], Step [400/700], Loss: 0.6427
Epoch [2/6], Step [600/700], Loss: 0.6291
Epoch [2/6], Step [700/700], Loss: 0.6201
Epoch [3/6], Step [200/700], Loss: 0.5336
Epoch [3/6], Step [400/700], Loss: 0.5281
Epoch [3/6], Step [600/700], Loss: 0.5239
Epoch [3/6], Step [700/700], Loss: 0.5266
Epoch [4/6], Step [200/700], Loss: 0.5961
Epoch [4/6], Step [400/700], Loss: 0.5586
Epoch [4/6], Step [600/700], Loss: 0.5333
Epoch [4/6], Step [700/700], Loss: 0.5200
Epoch [5/6], Step [200/700], Loss: 0.4068
Epoch [5/6], Step [400/700], Loss: 0.4123
Epoch [5/6], Step [600/700], Loss: 0.4095
Epoch [5/6], Step [700/700], Loss: 0.4069
Epoch [6/6], Step [200/700], Loss: 0.3628
Epoch [6/6], Step [400/700], Loss: 0.3656
Epoch [6/6], Step [600/700], Loss: 0.3633
Epoch [6/6], Step [700/700], Loss: 0.3608
Test Accuracy of the model: 83.40666666666667 %
```

8. lr = 0,05 and mom = 0,95

```
Epoch [1/6], Step [200/700], Loss: 0.6939
Epoch [1/6], Step [400/700], Loss: 0.6937
Epoch [1/6], Step [600/700], Loss: 0.6841
Epoch [1/6], Step [700/700], Loss: 0.6815
Epoch [2/6], Step [200/700], Loss: 0.6421
Epoch [2/6], Step [400/700], Loss: 0.6349
Epoch [2/6], Step [600/700], Loss: 0.6217
Epoch [2/6], Step [700/700], Loss: 0.6149
Epoch [3/6], Step [200/700], Loss: 0.5117
Epoch [3/6], Step [400/700], Loss: 0.5344
Epoch [3/6], Step [600/700], Loss: 0.5361
Epoch [3/6], Step [700/700], Loss: 0.5322
Epoch [4/6], Step [200/700], Loss: 0.4381
Epoch [4/6], Step [400/700], Loss: 0.4395
Epoch [4/6], Step [600/700], Loss: 0.4316
```

```
Epoch [4/6], Step [700/700], Loss: 0.4330
Epoch [5/6], Step [200/700], Loss: 0.3884
Epoch [5/6], Step [400/700], Loss: 0.3946
Epoch [5/6], Step [600/700], Loss: 0.3841
Epoch [5/6], Step [700/700], Loss: 0.3820
Epoch [6/6], Step [200/700], Loss: 0.3518
Epoch [6/6], Step [400/700], Loss: 0.3476
Epoch [6/6], Step [600/700], Loss: 0.3442
Epoch [6/6], Step [700/700], Loss: 0.3445
Test Accuracy of the model: 84.38666666666667 %
```