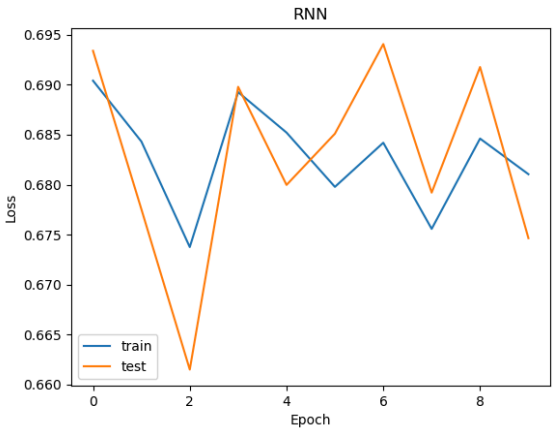
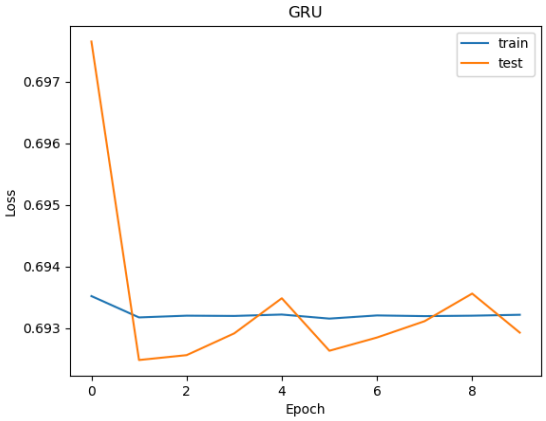
Ex3 idl
Amir Kelman, Omer Ben Haim

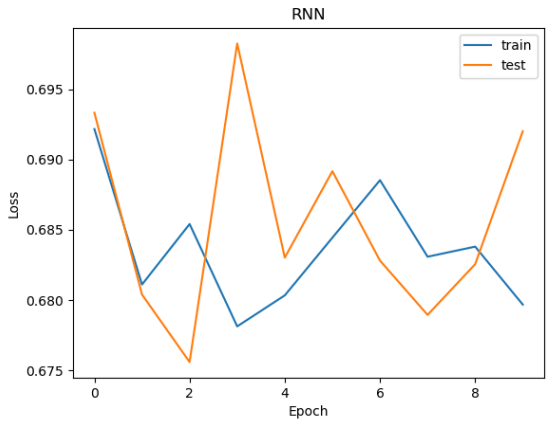
1. The architecture of RNN and GRU is a single FC layer that takes the hidden state and pass it through an MLP to get the two dimension classification as an output.

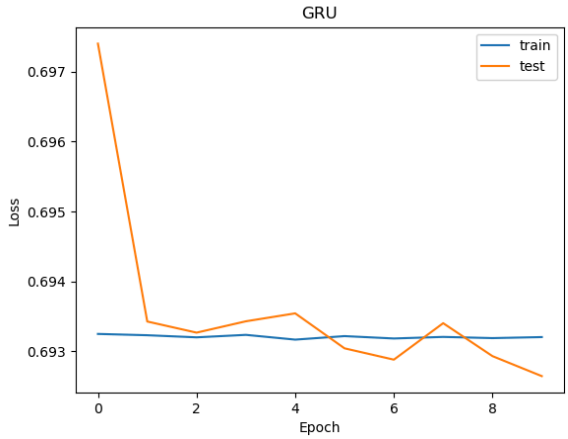
   Hidden state size 64:





Hidden state size 128:





RNN hidden state 128:
Review: The plot of the action movie was predictable, relying on tired tropes and cliches. However, the acting performances were surprisingly nuanced, particularly from the supporting cast. The fast-paced action sequences were well-choreographed and kept me engaged throughout the film. Overall, it wasn't a groundbreaking movie, but it was a fun and entertaining popcorn flick.
Prediction: tensor([[-0.1537, 0.6568]], grad_fn=<AddmmBackward0>)
True label: pos

RGU hidden state 128:Review: The plot of the action movie was predictable, relying on tired tropes and cliches. However, the acting performances were surprisingly nuanced, particularly from the supporting cast. The fast-paced action sequences were well-choreographed and kept me engaged throughout the film. Overall, it wasn't a groundbreaking movie, but it was a fun and entertaining popcorn flick.
Prediction: tensor([[-0.0374, -0.0309]], grad_fn=<AddmmBackward0>)
True label: pos

We dont see much difference between the models.

But we should see that the review might full an RNN but not GRU.
RNNs:

- RNNs process information sequentially, potentially getting stuck on the initial negative sentiment or the negative ending about the review.
- They may struggle to capture the shift in sentiment towards the acting and action sequences later in the review.

GRUs:

- GRUs have gating mechanisms that allow them to control information flow and potentially learn to focus on relevant details.
- In this case, the GRU might learn to pay attention to the positive aspects of the acting and action despite the initial negative comments about the plot.
- This could lead to a more balanced or even slightly positive overall sentiment prediction for the movie.

This is the reason why we expect to see different prediction from RNN and GRU in some reviews.

2. FN:

['ah', 'hitchcock', 'it', 'hard', 'to', 'find', 'bad', 'hitchcock', 'movie', 'until', 'he', 'lost', 'it', 'after', 'the', 'birds', 'and', 'saboteur', 'proves', 'the', 'point', 'having', 'admired', 'most', 'of', 'this', 'director', 'work', 'for', 'many', 'years', 'had', 'managed', 'to', 'skip', 'this', 'one', 'perhaps', 'from', 'lack', 'of', 'interest', 'in', 'priscilla', 'lane', 'and', 'robert', 'cummings', 'as', 'lead', 'actors', 'was', 'of', 'course', 'familiar', 'with', 'the', 'statue', 'of', 'liberty', 'climax', 'from', 'having', 'seen', 'it', 'repeatedly', 'in', 'film', 'retrospectives', 'but', 'wrongly', 'assumed', 'the', 'story', 'leading', 'up', 'to', 'it', 'might', 'not', 'hold', 'my', 'interest', 'was', 'wrong', 'the', 'suspenseful', 'plot', 'gets', 'cooking', 'right', 'off', 'the', 'bat', 'through', 'chance', 'encounter', 'between', 'the', 'bad']
Sub-scores:

```
[-9.6229494e-01  4.1039519e+00  1.5871029e+00  2.2638962e+00
 -6.2456119e-01 -6.6807067e-01 -3.1257940e+01  4.1039519e+00
 -2.2793653e+00  4.9908999e-01  1.7786388e+00 -1.1336322e+00
  1.5871029e+00  2.7644584e-01  1.4856553e-01 -3.1265134e-01
  2.9770634e+00 -5.9799333e+00 -2.7002795e+00  1.4856553e-01
 -1.0251434e+00 -1.1839801e-01  1.3450848e+01  6.3338261e+00
  1.1832678e+00  3.0774024e-01 -6.2999725e-01  8.5065472e-01
  4.7639936e-01  3.0249031e+00  1.5834504e+00 -2.8742063e+00
  4.5138264e-01 -6.2456119e-01 -2.4259133e+00  3.0774024e-01
  3.9636597e-01 -1.5789781e+00  2.0324305e-01 -8.7652874e+00
  1.1832678e+00  9.8950088e-01  1.1343237e+00  1.4628046e+00
  6.3572776e-01  2.9770634e+00  8.0695218e-01  1.7048495e+00
  4.4272029e-01 -5.1591051e-01 -2.3941278e+00 -2.0333283e-01
  1.1832678e+00  2.7961075e+00 -4.0673840e-01  4.0880021e-01
  1.4856553e-01  3.8258478e-02  1.1832678e+00  3.4582632e+00
  6.6838098e+00  2.0324305e-01 -1.1839801e-01 -4.7545052e-01
  1.5871029e+00  9.7544216e-02  1.1343237e+00 -1.4265540e-01
  7.3303676e+00  4.1980989e-02 -2.2660694e+00  8.4796250e-01
  1.4856553e-01  1.0546372e+00  1.6359110e+00 -1.5050708e-03
 -6.2456119e-01  1.5871029e+00 -4.4044385e+00 -3.4325335e+00
 -2.0531343e-01  5.1839347e+00  9.8950088e-01 -2.0333283e-01
 -2.2521931e+01  1.4856553e-01  5.7438245e+00 -7.2686238e+00
  7.4166119e-01  1.9310099e+00 -1.2795639e+00 -1.7447098e+00
```

1.4856553e-01 8.1475683e-02 7.4315596e-01 2.7375996e-01
4.4721575e+00 1.5967333e+00 1.4856553e-01 -3.1257940e+01]
[-3.8409773e-02 -4.9506607e+00 -2.3741057e+00 -3.2706375e+00
1.0283561e+00 8.7744904e-01 3.6473415e+01 -4.9506607e+00
3.3303549e+00 -8.9435852e-01 -2.4112408e+00 2.3139985e+00
-2.3741057e+00 -5.8389711e-01 -4.4303516e-01 2.0632996e-01
-3.4405396e+00 8.5103931e+00 3.3514612e+00 -4.4303516e-01
1.5931168e+00 1.3363020e-01 -1.6096329e+01 -7.2651753e+00
-1.2423599e+00 -4.7927105e-01 2.9161370e-01 -1.5196128e+00
-7.0135790e-01 -2.8575354e+00 -1.7464986e+00 3.5465696e+00
8.1120385e-03 1.0283561e+00 3.2378502e+00 -4.7927105e-01
-8.4345126e-01 2.3577182e+00 -1.3844559e-01 1.1543935e+01
-1.2423599e+00 -9.0035534e-01 -1.6849258e+00 -1.7810657e+00
1.6577755e-01 -3.4405396e+00 -8.8120592e-01 -1.3473003e+00
-9.0902925e-01 7.5905919e-01 3.1413991e+00 1.4714852e-01
-1.2423599e+00 -4.1568699e+00 2.8244874e-01 -9.1480410e-01
-4.4303516e-01 8.3540253e-02 -1.2423599e+00 -3.9174736e+00
-7.8720951e+00 -1.3844559e-01 1.3363020e-01 1.6860503e-01
-2.3741057e+00 -4.9833044e-01 -1.6849258e+00 3.1161657e-01
-8.3289967e+00 -2.4426374e-01 4.8577871e+00 -5.8950305e-01
-4.4303516e-01 -1.2345648e+00 -2.0247133e+00 2.4402605e-01
1.0283561e+00 -2.3741057e+00 5.7818298e+00 4.1038327e+00
2.6764995e-01 -5.5367703e+00 -9.0035534e-01 1.4714852e-01
2.6657396e+01 -4.4303516e-01 -7.2978725e+00 1.0148958e+01
-6.9387138e-01 -2.6372781e+00 1.4521773e+00 1.7562765e+00
-4.4303516e-01 2.6321519e-02 -1.0870900e+00 -7.2409309e-02
-5.1530151e+00 -1.8562517e+00 -4.4303516e-01 3.6473415e+01]
Final scores:
-0.30274948
0.3995228
Softmaxed predictions:
tensor([0.3313, 0.6687])
True labels:
1.0
0.0

This example is hard to classify according to the words written, so it is makes seans that the model got false negative.

FP:

['whoa', 'mean', 'whoa', 'mean', 'whoa', 'whoa', 'br', 'br', 'saw', 'this', 'movie', 'waaay', 'back', 'when', 'was', 'eight', 'in', 'back', 'then', 'cgi', 'films', 'were', 'rarity', 'and', 'good', 'ones', 'even', 'more', 'so', 'also', 'back', 'then', 'we', 'listened', 'to', 'things', 'called', 'cd', 'players', 'but', 'digress', 'used', 'to', 'like', 'this', 'movie', 'lot', 'way', 'back', 'then', 'and', 'up', 'till', 'viewing', 'it', 'again', 've', 'held', 'reaally', 'fond', 'memories', 'of', 'it', 'hey', 'it', 'don', 'bluth', 'anyone', 'who', 'hates', 'all', 'dogs', 'go', 'to', 'heaven', 'is', 'clearly', 'robot', 'but', 'again', 'digress', 'br', 'br', 'then', 'saw', 'it', 'again', 'this', 'really', 'isn', 'one', 'of', 'his', 'best', 'can', 'say', 'now', 'eleven', 'years', 'later']
Sub-scores:
[-7.137762 -9.987018 -7.137762 -9.987018 -7.137762 -7.137762
-1.9745932 -1.9745932 0.10854793 -0.03658469 -1.829081 0.38422942
-0.11262897 0.46949783 0.09027092 0.21143098 0.87213176 -0.11262897
0.3381421 -4.0885353 2.1610951 -2.861019 6.5893846 2.624957
9.270189 -4.4029045 -0.90401113 1.4313183 0.472198 0.5927407
-0.11262897 0.3381421 1.1471877 1.7552224 -0.66667366 0.27641323
0.12894225 3.250553 1.2912339 0.0280642 -1.9198546 -1.1993052
-0.66667366 0.05644881 -0.03658469 -1.829081 0.33751082 1.3560992
-0.11262897 0.3381421 2.624957 -0.06174014 -0.2369472 1.9413861
0.95287466 0.42792 3.4316926 0.84722924 0.38422942 8.46941
12.267283 1.0728744 0.95287466 -2.789643 0.95287466 0.25185138
2.879174 -3.5318272 0.53930634 -0.18092924 1.4581604 -0.09888142
-1.4286034 -0.66667366 2.4714496 1.004282 -1.5624988 -0.03614398
0.0280642 0.42792 -1.9198546 -1.9745932 -1.9745932 0.3381421

```
 0.10854793 0.95287466 0.42792  -0.03658469 3.6009593 -8.548371
 0.23183037 1.0728744  3.1315255  8.699642  -0.42262566 -3.1804893
 0.26656687 3.3939981  1.0804894  0.36967245]
[ 8.530176  12.25254   8.530176  12.25254   8.530176
  8.530176   3.141699  3.141699  -0.22929901 -0.2171026
  2.7380881 -0.31044263 -0.06290331 -0.5586591  -0.2598943
 -0.57987535 -1.4105446  -0.06290331 -0.5264284  4.851653
 -2.2527406  3.1164489  -7.729147  -3.2044876 -11.453356
  5.506509   0.9223876  -2.0892518  -1.042139  -1.3355848
 -0.06290331 -0.5264284  -1.9220096  -2.197886   1.0264475
 -0.56453836 -0.7160979  -3.8652337  -1.8219268  -0.39201882
  2.6530547  1.3227589  1.0264475  -0.47848475 -0.2171026
  2.7380881 -1.1204675  -1.7371857  -0.06290331 -0.5264284
 -3.2044876  0.1990692  0.06484229 -2.132728  -1.7058206
 -0.69492877 -4.1769123  -1.1132479  -0.31044263 -9.956885
-12.777362  -1.0916635  -1.7058206  3.08982   -1.7058206
 -0.3627028  -3.0874503  4.20612   -0.7837981  1.1047246
 -1.8682978  -0.24976541  1.547111   1.0264475  -2.6326077
 -1.7121542  1.4001184  -0.23027028 -0.39201882 -0.69492877
  2.6530547  3.141699  3.141699  -0.5264284  -0.22929901
 -1.7058206  -0.69492877 -0.2171026  -5.3097086  10.162999
 -0.6753706  -1.0916635  -3.1812923 -11.115254   0.6578224
  4.1687436  -0.85460687 -3.9041219  -1.3170595  -0.7731279 ]
Final scores:
0.009648603
-0.063469365
Softmaxed predictions:
tensor([0.5183, 0.4817])
True labels:
0.0
1.0
```

This example is hard to classify according to the words written also, so it got false positive.

TP:

```
['i', 'would', 'not', 'compare', 'it', 'to', 'le', 'placard', 'which', 'imho', 'had', 'more', 'comic', 'moments', 'but',
'romuald', 'juliette', 'while', 'being', 'slow', 'starter', 'certainly', 'kept', 'your', 'attention', 'going', 'throughout',
'the', 'film', 'nicely', 'paced', 'and', 'reaching', 'heart', 'warming', 'conclusion', 'there', 'were', 'many',
'marvellous', 'comedic', 'moments', 'some', 'brilliant', 'pathos', 'and', 'realistic', 'situation', 'acting', 'by', 'all',
'actors', 'br', 'br', 'it', 'was', 'typically', 'french', 'film', 'in', 'which', 'while', 'confronting', 'prejudices', 'and',
'phobias', 'which', 'in', 'turn', 'the', 'made', 'the', 'viewer', 'confront', 'his', 'own', 'shortcomings', 'am',
'certainly', 'pleased', 'to', 'have', 'this', 'in', 'my', 'library', 'and', 'will', 'no', 'doubt', 'watch', 'it', 'time', 'and',
'time', 'again', 'which', 'to', 'me', 'is']
Sub-scores:
[ 2.142517  -2.9067721  -3.7185957  1.77101    1.5263745
 -0.70478123  2.6362786  -4.712139   0.7117578  0.37590614
 -3.3101134  2.4310732  3.9518363  10.384288  -0.27395138
 -2.5949244  1.1707193  0.6130171  -1.8356205  -5.3180447
-12.525059   0.09686925 -0.73018414  4.073784   1.4424546
 -2.4871807  3.4980483  0.16386701  0.19815548 10.378152
  8.957759   2.8141072  0.6430884  0.66334337  0.407558
 -3.1425133  -1.7505192  -4.197142   2.81571   -7.7530127
  7.72905   10.384288  -1.5703787  14.3588    9.379092
  2.8141072  4.950253  -1.2096505  -2.8525305  0.5567498
  1.5988177  -2.6801095  -1.6201214  -1.6201214  1.5263745
 -0.92867696 -0.72404224  1.4318212  0.19815548  1.166563
  0.7117578  0.6130171  4.2128754  -0.9469474  2.8141072
  4.3231993  0.7117578  1.166563  -1.415792   0.16386701
 -2.1559591  0.16386701  1.0466164  2.3664305  1.9176869
  3.9254148 -13.188708  -1.856191   0.09686925  5.361753
 -0.70478123 -2.5405676  0.19065474  1.166563   4.433329
```

```
  1.505465   2.8141072   0.5342222  -5.467915   -2.005281
  0.9319909  1.5263745   1.000311    2.8141072  1.000311
  0.2302672  0.7117578  -0.70478123  1.1626222  1.314688 ]
[-2.7393293e+00  3.6887102e+00  4.5819674e+00 -2.0383544e+00
 -2.2313373e+00  1.1638277e+00 -2.7187352e+00  6.2066889e+00
 -1.2815914e+00 -2.8804353e-01  4.1459613e+00 -2.8252637e+00
 -4.6916871e+00 -1.1602149e+01  2.2541554e-01  3.7349916e+00
 -1.7104634e+00 -7.8426468e-01  1.8547727e+00  7.1067214e+00
  1.5423638e+01 -4.7830930e-03  8.7364590e-01 -4.1766214e+00
 -1.7358502e+00  2.9254005e+00 -3.9438498e+00 -3.3857948e-01
 -1.4066377e-01 -1.2706653e+01 -1.1518544e+01 -3.1991253e+00
 -8.5658997e-01 -1.1439855e+00 -5.3211147e-01  4.2098942e+00
  1.7320551e+00  5.0391607e+00 -2.6558018e+00  9.3151407e+00
 -8.5977697e+00 -1.1602149e+01  1.9824628e+00 -1.7383343e+01
 -1.0400074e+01 -3.1991253e+00 -6.2090960e+00  8.4851027e-01
  3.3049803e+00 -1.7163202e-01 -1.8431326e+00  3.4841437e+00
  2.5926728e+00  2.5926728e+00 -2.2313373e+00  9.6968406e-01
  7.0710397e-01 -1.7662762e+00 -1.4066377e-01 -1.6625907e+00
 -1.2815914e+00 -7.8426468e-01 -3.9075742e+00  2.0205085e+00
 -3.1991253e+00 -3.3427691e+00 -1.2815914e+00 -1.6625907e+00
  1.8442765e+00 -3.3857948e-01  2.1989777e+00 -3.3857948e-01
 -1.1146857e+00 -2.5936103e+00 -1.6085973e+00 -3.8771400e+00
  1.6375116e+01  1.8542457e+00 -4.7830930e-03 -6.7569351e+00
  1.1638277e+00  2.9132597e+00 -1.0123440e-01 -1.6625907e+00
 -4.6382470e+00 -1.5594114e+00 -3.1991253e+00 -9.1636080e-01
  6.7112832e+00  2.7786405e+00 -1.4353147e+00 -2.2313373e+00
 -1.1298600e+00 -3.1991253e+00 -1.1298600e+00 -4.8668906e-01
 -1.2815914e+00  1.1638277e+00 -1.1069103e+00 -2.0578785e+00]
Final scores:
0.68741196
-0.71565336
Softmaxed predictions:
tensor([0.8027, 0.1973])
True labels:
1.0
0.0
```

I assume that the positive words gave that review a good indication to predict it positive.

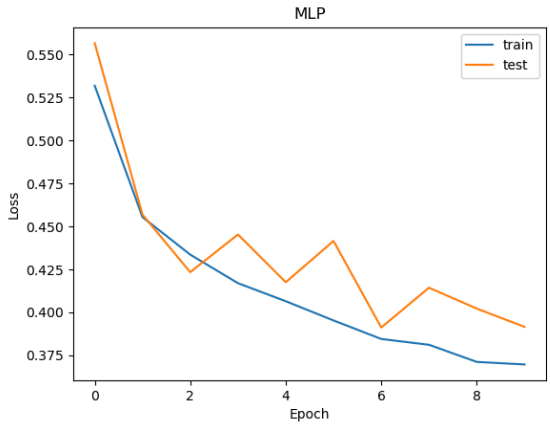
TN:

```
['maybe', 'this', 'was', 'an', 'important', 'movie', 'and', 'that', 'why', 'people', 'rank', 'it', 'so', 'highly', 'but',
'honestly', 'it', 'isn', 'very', 'good', 'in', 'hindsight', 'it', 'easy', 'to', 'see', 'that', 'chaplin', 'probably', 'all', 'of',
'hollywood', 'was', 'incredibly', 'naive', 'about', 'the', 'magnitude', 'of', 'what', 'was', 'really', 'going', 'on', 'in',
'the', 'ghettos', 'so', 'you', 'can', 'fault', 'him', 'too', 'much', 'for', 'the', 'disconnect', 'that', 'affects', 'modern',
'viewer', 'but', 'the', 'disconnect', 'remains', 'br', 'br', 'more', 'disappointingly', 'the', 'movie', 'is', 'just', 'clunky',
'it', 'as', 'if', 'chaplin', 'had', 'no', 'idea', 'that', 'movies', 'had', 'progressed', 'in', 'sophistication', 'since', 'the',
'silent', 'era', 'the', 'set', 'pieces', 'those', 'involving', 'both', 'the', 'jewish', 'barber']
Sub-scores:
[ -3.7924578   0.3830603  -0.7525213  -0.13662228  8.184943
  -2.4552605   3.094359   -0.23695911 -1.4548186   1.5784135
   0.9707887   1.6638821   0.60419774  7.964793   -0.05242509
  -5.963566    1.6638821  -6.6093264   8.27065    10.43884
   1.3565571 -10.813774    1.6638821   7.8843346  -0.73485315
   0.26012394 -0.23695911  4.721201   -5.7880793   1.7715031
   1.1889523  -0.7465397  -0.7525213  10.928817  -15.235921
   0.10272518  0.22272083  0.6926577   1.1889523   0.43013242
  -0.7525213   3.731319   -2.4855657   0.87523484  1.3565571
   0.22272083 -0.13031314  0.60419774  1.667081   -0.26673433
 -11.077882   -0.11041825 -7.136673    0.04439213  0.5712068
   0.22272083  0.56720525 -0.23695911  3.0977397   5.117733
   0.9283573  -0.05242509  0.22272083  0.56720525  0.06361818
  -1.3088187  -1.3088187   2.8835192 -26.202286    0.22272083
  -2.4552605   1.5032771  -2.798163  -27.315264    1.6638821
```

```
  0.52667785 -1.9166088   4.721201   -3.2973094  -5.7999296
 -1.6424457  -0.23695911  1.0850643  -3.2973094   1.9369503
  1.3565571   6.8115606   1.2671019   0.22272083  0.92646015
  4.165094    0.22272083  0.6424461   1.5075455   0.37448272
  1.8711005   1.9134824   0.22272083  2.4077044   0.34572572]
[ 4.69914913e+00 -3.69002044e-01  7.57477760e-01 -7.35548958e-02
 -9.52826023e+00  3.45853090e+00 -3.47372246e+00 -1.77880555e-01
  2.07307124e+00 -1.18972015e+00 -9.24845695e-01 -2.32299018e+00
 -8.64171624e-01 -9.18945599e+00 -3.13202962e-02  6.07441807e+00
 -2.32299018e+00  8.01616001e+00 -1.03315992e+01 -1.26317968e+01
 -1.85003865e+00  1.25022745e+01 -2.32299018e+00 -1.00455503e+01
  1.27147388e+00 -2.40042850e-01 -1.77880555e-01 -4.88071442e+00
  7.55103207e+00 -1.99142206e+00 -1.21914339e+00  1.14829361e+00
  7.57477760e-01 -1.40360880e+01  1.74545956e+01  5.40005006e-02
 -3.86578470e-01  1.23015635e-01 -1.21914339e+00 -7.26055086e-01
  7.57477760e-01 -5.04359913e+00  2.91702318e+00 -1.17801356e+00
 -1.85003865e+00 -3.86578470e-01  1.05678689e+00 -8.64171624e-01
 -1.56789029e+00  7.27077186e-01  1.43966475e+01  6.25536323e-01
  8.65668583e+00  3.66414636e-01 -7.40692675e-01 -3.86578470e-01
  5.98615825e-01 -1.77880555e-01 -2.47473001e+00 -5.34328032e+00
 -9.83068585e-01 -3.13202962e-02 -3.86578470e-01  5.98615825e-01
 -1.37769297e-01  2.25506926e+00  2.25506926e+00 -3.34007931e+00
  3.11076450e+01 -3.86578470e-01  3.45853090e+00 -2.25574684e+00
  2.94599509e+00  3.18920288e+01 -2.32299018e+00 -9.52411234e-01
  2.27299142e+00 -4.88071442e+00  4.10329866e+00  7.06238508e+00
  2.08145714e+00 -1.77880555e-01 -1.09743607e+00  4.10329866e+00
 -2.57996750e+00 -1.85003865e+00 -8.23536110e+00 -1.52129066e+00
 -3.86578470e-01 -1.03642499e+00 -4.01302576e+00 -3.86578470e-01
 -9.87787247e-01 -1.46965349e+00 -3.04379970e-01 -1.75188684e+00
 -2.19958973e+00 -3.86578470e-01 -2.46763039e+00 -2.45108351e-01]
Final scores:
-0.19700173
0.30864757
Softmaxed predictions:
tensor([0.3762, 0.6238])
True labels:
0.0
1.0
```

'Disappointingly' and 'clunky' did the work fine.

The MLP loss:



3. The code is attached and can be found under the ExLRestSelfAtten class.
4. The experiment was conducted again and we looked at the same 4 examples we presented in question 2. We can see the results have improved, (add the details when we get them) we can explain the improvement with the fact that the model used in question 4 was a transformer, which takes into account the score that the transformer gives to every combination of words in the word;s window - or in other words - it “looks” at the word’s surrounding (in our case - a five words window). This gives the model a better ability to predict the sentences meaning, since the MLP in question 2 does not take into account the word’s “window” - it performs poorly compared to the transformer.