# ECGR 5106 Homework 5: Transformers

Lauren Bourque (Student ID: 801264851)

April 26, 2024

## GitHub Link

Click here to view the code

## Problem 1.

**Sequences of 10:**
The model structure can be seen in Figure 1 below, providing the computational complexity such as number of parameters for each layer:

Figure 1: Model Structure

```
===================================================================================================
Layer (type:depth-idx)                      Output Shape              Param #
===================================================================================================
CharTransformer                             [1900, 44]                --
├─Embedding: 1-1                            [1900, 10, 128]           5,632
├─TransformerEncoder: 1-2                   [1900, 10, 128]           --
│    └─ModuleList: 2-1                      --                        --
│    │    └─TransformerEncoderLayer: 3-1    [1900, 10, 128]           593,024
│    │    └─TransformerEncoderLayer: 3-2    [1900, 10, 128]           593,024
│    │    └─TransformerEncoderLayer: 3-3    [1900, 10, 128]           593,024
├─Linear: 1-3                               [1900, 44]                5,676
===================================================================================================
Total params: 1,790,380
Trainable params: 1,790,380
Non-trainable params: 0
Total mult-adds (G): 3.03
===================================================================================================
Input size (MB): 0.15
Forward/backward pass size (MB): 1129.12
Params size (MB): 6.37
Estimated Total Size (MB): 1135.64
===================================================================================================
```

The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 2 below:

Figure 2: Training Loss + Accuracy

```
Epoch 5, Loss: 2.829401969909668, Validation Loss: 2.6700103282928467, Validation Accuracy: 0.27941176295280457, Execution Ti
me: 18.270447731018066 seconds
Epoch 10, Loss: 2.552777051925659, Validation Loss: 2.523683786392212, Validation Accuracy: 0.23529411852359772, Execution Ti
me: 16.170519828796387 seconds
Epoch 15, Loss: 2.440364122390747, Validation Loss: 2.4505257606506348, Validation Accuracy: 0.27941176295280457, Execution T
ime: 16.71851110458374 seconds
Epoch 20, Loss: 2.3740100860595703, Validation Loss: 2.412954807281494, Validation Accuracy: 0.2710084021091461, Execution Ti
me: 16.37756371498108 seconds
Epoch 25, Loss: 2.321685791015625, Validation Loss: 2.368964910507202, Validation Accuracy: 0.29201680421829224, Execution Ti
me: 17.226407289505005 seconds
Epoch 30, Loss: 2.2885217666625977, Validation Loss: 2.3425052165985107, Validation Accuracy: 0.2899159789085388, Execution T
ime: 16.2433762550354 seconds
Epoch 35, Loss: 2.2667012214660645, Validation Loss: 2.328082799911499, Validation Accuracy: 0.2836134433746338, Execution Ti
me: 16.507776260375977 seconds
Epoch 40, Loss: 2.246110200881958, Validation Loss: 2.3244121074676514, Validation Accuracy: 0.29201680421829224, Execution T
ime: 16.902719259262085 seconds
Epoch 45, Loss: 2.23183012008667, Validation Loss: 2.316819667816162, Validation Accuracy: 0.3025210201740265, Execution Tim
e: 16.58337688446045 seconds
Epoch 50, Loss: 2.2328975200653076, Validation Loss: 2.319289445877075, Validation Accuracy: 0.28151261806488037, Execution T
ime: 17.38715934753418 seconds
Total Execution Time: 834.1697237491608 seconds
```

**Sequences of 20:**

The model structure can be seen in Figure 3 below, providing the computational complexity such as number of parameters for each layer:

Figure 3: Model Structure

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
CharTransformer                          [1892, 44]                --
├─Embedding: 1-1                         [1892, 20, 128]           5,632
├─TransformerEncoder: 1-2                [1892, 20, 128]           --
│    └─ModuleList: 2-1                   --                        --
│    │    └─TransformerEncoderLayer: 3-1 [1892, 20, 128]           593,024
│    │    └─TransformerEncoderLayer: 3-2 [1892, 20, 128]           593,024
│    │    └─TransformerEncoderLayer: 3-3 [1892, 20, 128]           593,024
├─Linear: 1-3                            [1892, 44]                5,676
==========================================================================================
Total params: 1,790,380
Trainable params: 1,790,380
Non-trainable params: 0
Total mult-adds (G): 3.01
==========================================================================================
Input size (MB): 0.30
Forward/backward pass size (MB): 2248.06
Params size (MB): 6.37
Estimated Total Size (MB): 2254.73
==========================================================================================
```

The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 4 below:

Figure 4: Training Loss + Accuracy

```
Epoch 5, Loss: 2.902303695678711, Validation Loss: 2.8105857372283936, Validation Accuracy: 0.20464135706424713, Execution Ti
me: 33.94730591773987 seconds
Epoch 10, Loss: 2.629087448120117, Validation Loss: 2.5793728828430176, Validation Accuracy: 0.2405063360929489, Execution Ti
me: 33.82742261886597 seconds
Epoch 15, Loss: 2.49463152885437, Validation Loss: 2.50187349319458, Validation Accuracy: 0.24261602759361267, Execution Tim
e: 33.6160569190979 seconds
Epoch 20, Loss: 2.4157893657684326, Validation Loss: 2.4567251205444336, Validation Accuracy: 0.27426159381866455, Execution
Time: 33.36484980583191 seconds
Epoch 25, Loss: 2.3571343421936035, Validation Loss: 2.4352853298187256, Validation Accuracy: 0.25738397240638733, Execution
Time: 33.260478019714355 seconds
Epoch 30, Loss: 2.320802688598633, Validation Loss: 2.3991174697875977, Validation Accuracy: 0.2637130916118622, Execution Ti
me: 33.400601863861084 seconds
Epoch 35, Loss: 2.2873716354370117, Validation Loss: 2.372976541519165, Validation Accuracy: 0.27848100662231445, Execution T
ime: 32.386797189712524 seconds
Epoch 40, Loss: 2.270092725753784, Validation Loss: 2.3724288940429688, Validation Accuracy: 0.2552742660045624, Execution Ti
me: 32.659749269485474 seconds
Epoch 45, Loss: 2.2507073879241943, Validation Loss: 2.361520290374756, Validation Accuracy: 0.2805907130241394, Execution Ti
me: 33.32268929481506 seconds
Epoch 50, Loss: 2.2436301708221436, Validation Loss: 2.3631558418273926, Validation Accuracy: 0.2763713002204895, Execution T
ime: 32.71213507652283 seconds
Total Execution Time: 1658.8146982192993 seconds
```

**Sequences of 30:**

The model structure can be seen in Figure 5 below, providing the computational complexity such as number of parameters for each layer:

Figure 5: Model Structure

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
CharTransformer                          [1884, 44]                --
├─Embedding: 1-1                         [1884, 30, 128]           5,632
├─TransformerEncoder: 1-2                [1884, 30, 128]           --
│    └─ModuleList: 2-1                   --                        --
│    │    └─TransformerEncoderLayer: 3-1 [1884, 30, 128]           593,024
│    │    └─TransformerEncoderLayer: 3-2 [1884, 30, 128]           593,024
├─Linear: 1-3                            [1884, 44]                5,676
==========================================================================================
Total params: 1,197,356
Trainable params: 1,197,356
Non-trainable params: 0
Total mult-adds (G): 2.01
==========================================================================================
Input size (MB): 0.45
Forward/backward pass size (MB): 2257.85
Params size (MB): 4.26
Estimated Total Size (MB): 2262.56
==========================================================================================
```

The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 6 below:

Figure 6: Training Loss + Accuracy

```
Epoch 5, Loss: 2.7724125385284424, Validation Loss: 2.739386796951294, Validation Accuracy: 0.22669491171836853, Execution Ti
me: 34.22837591171265 seconds
Epoch 10, Loss: 2.542022705078125, Validation Loss: 2.5789334774017334, Validation Accuracy: 0.23516948521137238, Execution T
ime: 34.438804388046265 seconds
Epoch 15, Loss: 2.441145420074463, Validation Loss: 2.5189452171325684, Validation Accuracy: 0.22669491171836853, Execution T
ime: 34.14780068397522 seconds
Epoch 20, Loss: 2.3785905838012695, Validation Loss: 2.4940552711486816, Validation Accuracy: 0.21822033822536469, Execution
Time: 33.687227964401245 seconds
Epoch 25, Loss: 2.3335747718811035, Validation Loss: 2.459984302520752, Validation Accuracy: 0.24788135290145874, Execution T
ime: 34.31617975234985 seconds
Epoch 30, Loss: 2.3007843494415283, Validation Loss: 2.446418523788452, Validation Accuracy: 0.24788135290145874, Execution T
ime: 33.46302628517151 seconds
Epoch 35, Loss: 2.271259307861328, Validation Loss: 2.438570499420166, Validation Accuracy: 0.25, Execution Time: 34.39435172
080994 seconds
Epoch 40, Loss: 2.24881911277771, Validation Loss: 2.4344990253448486, Validation Accuracy: 0.24788135290145874, Execution Ti
me: 34.640878200531006 seconds
Epoch 45, Loss: 2.238251209259033, Validation Loss: 2.4346768856048584, Validation Accuracy: 0.24364407360553741, Execution T
ime: 34.25514793395996 seconds
Epoch 50, Loss: 2.218820810317993, Validation Loss: 2.438627004623413, Validation Accuracy: 0.25211864709854126, Execution Ti
me: 33.24434781074524 seconds
Total Execution Time: 1713.871531009674 seconds
```

Overall, the sequences of 10 provided the best validation accuracy.

# Problem 2.

**Sequences of 20:**
The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 7 below:

Figure 7: Training Loss + Accuracy

```
Epoch 1, Train Loss: 2.511201670412319, Validation Loss: 2.488522491706801, Validation Accuracy: 0.2612519724573232
Epoch 2, Train Loss: 2.482445878528348, Validation Loss: 2.4822470452655185, Validation Accuracy: 0.2686128245588868
Epoch 3, Train Loss: 2.4770272818857997, Validation Loss: 2.478883504696546, Validation Accuracy: 0.2658334528762014
Epoch 4, Train Loss: 2.4738775046240087, Validation Loss: 2.4746318428130936, Validation Accuracy: 0.2690924903170277
Epoch 5, Train Loss: 2.471594839454501, Validation Loss: 2.478012214208364, Validation Accuracy: 0.2635830583847368
Epoch 6, Train Loss: 2.4701393740263686, Validation Loss: 2.472646136346121, Validation Accuracy: 0.26820936737914214
Epoch 7, Train Loss: 2.4691859109383576, Validation Loss: 2.4724686447097586, Validation Accuracy: 0.26885489886673364
Epoch 8, Train Loss: 2.468294793710256, Validation Loss: 2.474320502965903, Validation Accuracy: 0.26847385597475254
Epoch 9, Train Loss: 2.4675045568175884, Validation Loss: 2.471005383394247, Validation Accuracy: 0.2692673217615837
Epoch 10, Train Loss: 2.46693564340282, Validation Loss: 2.4720575176751214, Validation Accuracy: 0.2689580045904461
Epoch 11, Train Loss: 2.466498386396464, Validation Loss: 2.4715782235845607, Validation Accuracy: 0.2688593817242863
Epoch 12, Train Loss: 2.465959999794222, Validation Loss: 2.471158846227361, Validation Accuracy: 0.2685679959833596
Epoch 13, Train Loss: 2.46548342567967, Validation Loss: 2.46966001108821, Validation Accuracy: 0.26867110170707215
Epoch 14, Train Loss: 2.4651649310862105, Validation Loss: 2.4685175512223823, Validation Accuracy: 0.2689804188782097
Epoch 15, Train Loss: 2.464841829073376, Validation Loss: 2.4692465988126893, Validation Accuracy: 0.26932559890976904
Epoch 16, Train Loss: 2.4646516742673463, Validation Loss: 2.4690229703254465, Validation Accuracy: 0.27023113613541816
Epoch 17, Train Loss: 2.464382179160625, Validation Loss: 2.468950020601718, Validation Accuracy: 0.2684648902596471
Epoch 18, Train Loss: 2.4640671391515165, Validation Loss: 2.4699773849330797, Validation Accuracy: 0.268451441686989
Epoch 19, Train Loss: 2.4636682573443522, Validation Loss: 2.468704862516625, Validation Accuracy: 0.26930766747955814
Epoch 20, Train Loss: 2.4635618337342016, Validation Loss: 2.468592381036157, Validation Accuracy: 0.26942422177592884

Results for sequence length: 20

Transformer Model:
Training Loss: 2.4635618337342016
Validation Loss: 2.468592381036157
Validation Accuracy: 0.26942422177592884
Execution Time: 1314.9288251399994 seconds
```

**Sequences of 30:**
The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 8 below:

Figure 8: Training Loss + Accuracy

```
Epoch 1, Train Loss: 2.512194059487107, Validation Loss: 2.4756848642289833, Validation Accuracy: 0.2676000143454014
Epoch 2, Train Loss: 2.4838817431735323, Validation Loss: 2.4719237631709325, Validation Accuracy: 0.26871178295407677
Epoch 3, Train Loss: 2.478510127512279, Validation Loss: 2.4683206546966496, Validation Accuracy: 0.2675775996557104
Epoch 4, Train Loss: 2.4751962940781014, Validation Loss: 2.4687589308614744, Validation Accuracy: 0.2681738304014919
Epoch 5, Train Loss: 2.473546355335893, Validation Loss: 2.4682709120912323, Validation Accuracy: 0.26773450248354763
Epoch 6, Train Loss: 2.4718424746413343, Validation Loss: 2.4671372768473514, Validation Accuracy: 0.2689135151612961
Epoch 7, Train Loss: 2.470636532797343, Validation Loss: 2.464689446360919, Validation Accuracy: 0.26831728441551456
Epoch 8, Train Loss: 2.4699963924009096, Validation Loss: 2.4620981746349484, Validation Accuracy: 0.2706035827640002
Epoch 9, Train Loss: 2.4690966822511187, Validation Loss: 2.464325696504086, Validation Accuracy: 0.2700880449011064
Epoch 10, Train Loss: 2.468264890605929, Validation Loss: 2.465640468090537, Validation Accuracy: 0.27017322072193234
Epoch 11, Train Loss: 2.4679549866790444, Validation Loss: 2.4628491697050268, Validation Accuracy: 0.2688059246507791
Epoch 12, Train Loss: 2.4673739529063012, Validation Loss: 2.4621495505061777, Validation Accuracy: 0.27048702637760685
Epoch 13, Train Loss: 2.466970974379752, Validation Loss: 2.4636309491381776, Validation Accuracy: 0.2695456094105833
Epoch 14, Train Loss: 2.4667180668880824, Validation Loss: 2.462107895570196, Validation Accuracy: 0.2680841716427278
Epoch 15, Train Loss: 2.4664919807405283, Validation Loss: 2.4625976337561326, Validation Accuracy: 0.26871626589201497
Epoch 16, Train Loss: 2.466095617554658, Validation Loss: 2.4621888012276996, Validation Accuracy: 0.26898524216830744
Epoch 17, Train Loss: 2.465913338955930, Validation Loss: 2.4618011659907815, Validation Accuracy: 0.27005666433553893
Epoch 18, Train Loss: 2.4654697663647176, Validation Loss: 2.461008305828967, Validation Accuracy: 0.2690121397959367
Epoch 19, Train Loss: 2.4652400905483542, Validation Loss: 2.460980305145065, Validation Accuracy: 0.26958595585202716
Epoch 20, Train Loss: 2.4649602108898487, Validation Loss: 2.460609679763612, Validation Accuracy: 0.2689090322233579

Results for sequence length: 20

Transformer Model:
Training Loss: 2.4649602108898487
Validation Loss: 2.460609679763612
Validation Accuracy: 0.2689090322233579
Execution Time: 1632.793774843216 seconds
```

**Sequences of 50:**
The training loss, validation loss, validation accuracy, and total execution time can be seen in Figure 9 below:

Figure 9: Training Loss + Accuracy

```
Training Transformer model...
Epoch 1, Train Loss: 2.511785074914826, Validation Loss: 2.485132347520965, Validation Accuracy: 0.26712095400340713
Epoch 2, Train Loss: 2.482837775280943, Validation Loss: 2.477877752908963, Validation Accuracy: 0.2683807047431184
Epoch 3, Train Loss: 2.477420491778048, Validation Loss: 2.472766340489688, Validation Accuracy: 0.26829104276876176
Epoch 4, Train Loss: 2.4744677833855384, Validation Loss: 2.471256213909434, Validation Accuracy: 0.26762306105980455
Epoch 5, Train Loss: 2.4725065900425864, Validation Loss: 2.472710705158326, Validation Accuracy: 0.2633551510804268
Epoch 6, Train Loss: 2.47109892384777, Validation Loss: 2.4690889123629147, Validation Accuracy: 0.2683762216444006
Epoch 7, Train Loss: 2.4697847041657908, Validation Loss: 2.470059797969686, Validation Accuracy: 0.26894109208284767
Epoch 8, Train Loss: 2.469238147296938, Validation Loss: 2.4687171053931305, Validation Accuracy: 0.2692593920918139
Epoch 9, Train Loss: 2.4685092895840954, Validation Loss: 2.467085542450526, Validation Accuracy: 0.26848829911234645
Epoch 10, Train Loss: 2.467590626889597, Validation Loss: 2.466101540140771, Validation Accuracy: 0.26856451179054963
Epoch 11, Train Loss: 2.467185805708178, Validation Loss: 2.467029020286703, Validation Accuracy: 0.26823724558414774
Epoch 12, Train Loss: 2.466691150950026, Validation Loss: 2.4674211801781865, Validation Accuracy: 0.2689769568725903
Epoch 13, Train Loss: 2.466304288470113, Validation Loss: 2.4671611645697777, Validation Accuracy: 0.2682596610777369
Epoch 14, Train Loss: 2.466034688353592, Validation Loss: 2.466570901579999, Validation Accuracy: 0.2692907737828387
Epoch 15, Train Loss: 2.465792534719676, Validation Loss: 2.4653102789162027, Validation Accuracy: 0.269254908993096
Epoch 16, Train Loss: 2.4653771839451384, Validation Loss: 2.465447485569848, Validation Accuracy: 0.26928180758540304
Epoch 17, Train Loss: 2.4649988132624907, Validation Loss: 2.465517826930275, Validation Accuracy: 0.26895005828028334
Epoch 18, Train Loss: 2.464936914784293, Validation Loss: 2.4659160465058174, Validation Accuracy: 0.26860485967901016
Epoch 19, Train Loss: 2.464654972618881, Validation Loss: 2.4663554883479946, Validation Accuracy: 0.2708598583340805
Epoch 20, Train Loss: 2.464308036351689, Validation Loss: 2.464719433867, Validation Accuracy: 0.26953734421231956

Results for sequence length: 20

Transformer Model:
Training Loss: 2.464308036351689
Validation Loss: 2.464719433867
Validation Accuracy: 0.26953734421231956
Execution Time: 2449.5143270492554 seconds
```

Overall, the accuracy for all transformer-based models with sequences of 20, 30, and 50 were worse than the models with GRUs or LSTMs.

# Problem 3.

The training loss and accuracy can be seen in Figure 10 below:

Figure 10: Training Loss + Accuracy

```
Epoch [1/20], Loss: 5.9164, Training Accuracy: 0.0160
Epoch [2/20], Loss: 5.0685, Training Accuracy: 0.0427
Epoch [3/20], Loss: 4.5625, Training Accuracy: 0.0960
Epoch [4/20], Loss: 4.1173, Training Accuracy: 0.1573
Epoch [5/20], Loss: 3.7107, Training Accuracy: 0.2427
Epoch [6/20], Loss: 3.3208, Training Accuracy: 0.3067
Epoch [7/20], Loss: 2.9846, Training Accuracy: 0.4080
Epoch [8/20], Loss: 2.6569, Training Accuracy: 0.4373
Epoch [9/20], Loss: 2.3210, Training Accuracy: 0.5040
Epoch [10/20], Loss: 2.0154, Training Accuracy: 0.5333
Epoch [11/20], Loss: 1.7766, Training Accuracy: 0.5760
Epoch [12/20], Loss: 1.5314, Training Accuracy: 0.6240
Epoch [13/20], Loss: 1.2915, Training Accuracy: 0.6853
Epoch [14/20], Loss: 1.1415, Training Accuracy: 0.7013
Epoch [15/20], Loss: 0.9778, Training Accuracy: 0.7173
Epoch [16/20], Loss: 0.8153, Training Accuracy: 0.7413
Epoch [17/20], Loss: 0.6671, Training Accuracy: 0.7680
Epoch [18/20], Loss: 0.5718, Training Accuracy: 0.7733
Epoch [19/20], Loss: 0.4724, Training Accuracy: 0.7893
Epoch [20/20], Loss: 0.3980, Training Accuracy: 0.7920
```

The validation loss and accuracy can be seen in Figure 11 below:

Figure 11: Training Loss + Accuracy

```
Evaluation Loss: 0.1585, Evaluation Accuracy: 0.7947
```

The training and validation loss were lower in the transformer-based model than the RNN model without attention. However, the RNN model with attention had a better training and validation loss as well as validation accuracy.

# Problem 4.

The training loss and accuracy can be seen in Figure 12 below:

Figure 12: Training Loss + Accuracy

```
Epoch [1/20], Loss: 6.0092, Training Accuracy: 0.0107
Epoch [2/20], Loss: 5.1089, Training Accuracy: 0.0773
Epoch [3/20], Loss: 4.5889, Training Accuracy: 0.0987
Epoch [4/20], Loss: 4.1362, Training Accuracy: 0.1573
Epoch [5/20], Loss: 3.7471, Training Accuracy: 0.2347
Epoch [6/20], Loss: 3.3494, Training Accuracy: 0.3013
Epoch [7/20], Loss: 2.9824, Training Accuracy: 0.3493
Epoch [8/20], Loss: 2.6639, Training Accuracy: 0.4133
Epoch [9/20], Loss: 2.3408, Training Accuracy: 0.4907
Epoch [10/20], Loss: 2.0648, Training Accuracy: 0.5467
Epoch [11/20], Loss: 1.7792, Training Accuracy: 0.5893
Epoch [12/20], Loss: 1.5327, Training Accuracy: 0.6160
Epoch [13/20], Loss: 1.3426, Training Accuracy: 0.6773
Epoch [14/20], Loss: 1.1385, Training Accuracy: 0.6773
Epoch [15/20], Loss: 0.9598, Training Accuracy: 0.7280
Epoch [16/20], Loss: 0.8171, Training Accuracy: 0.7360
Epoch [17/20], Loss: 0.6916, Training Accuracy: 0.7573
Epoch [18/20], Loss: 0.5820, Training Accuracy: 0.7707
Epoch [19/20], Loss: 0.4867, Training Accuracy: 0.7760
Epoch [20/20], Loss: 0.4067, Training Accuracy: 0.7787
```

The validation loss and accuracy can be seen in Figure 13 below:

Figure 13: Training Loss + Accuracy

```
Evaluation Loss: 0.1699, Evaluation Accuracy: 0.7920
```

The RNN model with or without attention had a better training and validation loss as well as validation accuracy than the transformer-based model.