***1.***

* How many trainable parameters are in the model? It is printed in the terminal.
  + There are 346373 parameters in this model
* What is the best training accuracy?
  + Best training Accuracy: 75.768 percent
* What is the best validation accuracy? Is it better than the best training accuracies?
  + Best validation accuracy is 76.2800 percent
  + Yes
* According to the training and validation accuracies, does the model overfit your training data?
  + Based on the comparison between the training and validation accuracies, it appears that the model does not significantly overfit the training data. The validation accuracy is close to the training accuracy, indicating that the model generalizes well to unseen data.

There are 346373 parameters in this model

Use basic transformer for training

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 93.11it/s]

train Accuracy: 0.55528

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 144.64it/s]

valid Accuracy: 0.662

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 154.13it/s]

train Accuracy: 0.66504

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 184.70it/s]

valid Accuracy: 0.7084

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 153.94it/s]

train Accuracy: 0.692

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 183.42it/s]

valid Accuracy: 0.7348

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 155.22it/s]

train Accuracy: 0.70536

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 181.95it/s]

valid Accuracy: 0.714

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 155.93it/s]

train Accuracy: 0.72096

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 186.32it/s]

valid Accuracy: 0.7508

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 154.78it/s]

train Accuracy: 0.72992

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 188.16it/s]

valid Accuracy: 0.7284

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 154.21it/s]

train Accuracy: 0.7408

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 183.79it/s]

valid Accuracy: 0.748

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 155.15it/s]

train Accuracy: 0.74976

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 185.67it/s]

valid Accuracy: 0.746

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 154.50it/s]

train Accuracy: 0.75032

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 182.39it/s]

valid Accuracy: 0.7556

100%|████████████████████████████████████████████████████████████████| 196/196 [00:01<00:00, 155.43it/s]

train Accuracy: 0.75768

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 159.07it/s]

valid Accuracy: 0.7628

------------------------------------------------------------

best validation accuracy is 76.2800 percent

***2.***

* How many trainable parameters are in the model? It is printed in the terminal.
* There are 8069 parameters in this model
* What is the best training accuracy?
  + Best training accuracy is 52.5520 percent
* What is the best validation accuracy? Is it better than the best training accuracies?
  + Best validation accuracy is 53.3200 percent
  + Yes
* According to the training and validation accuracies, does the model overfit your training data?
  + Based on the comparison between the training and validation accuracies, it appears that the model does not significantly overfit the training data. The validation accuracy is slightly higher than the training accuracy, suggesting that the model generalizes well to unseen data.

There are 8069 parameters in this model

Use basic transformer for training

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:03<00:00, 51.21it/s]

train Accuracy: 0.4012

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 111.57it/s]

valid Accuracy: 0.4904

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 87.00it/s]

train Accuracy: 0.46664

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 141.51it/s]

valid Accuracy: 0.4844

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 86.89it/s]

train Accuracy: 0.4852

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 142.20it/s]

valid Accuracy: 0.5124

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 83.26it/s]

train Accuracy: 0.49712

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 139.18it/s]

valid Accuracy: 0.51

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 85.78it/s]

train Accuracy: 0.50488

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 142.06it/s]

valid Accuracy: 0.5216

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 85.42it/s]

train Accuracy: 0.5128

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 141.32it/s]

valid Accuracy: 0.5276

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 87.81it/s]

train Accuracy: 0.5144

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 141.79it/s]

valid Accuracy: 0.5292

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 83.40it/s]

train Accuracy: 0.51944

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 142.15it/s]

valid Accuracy: 0.528

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 88.26it/s]

train Accuracy: 0.52272

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 143.59it/s]

valid Accuracy: 0.5292

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 87.33it/s]

train Accuracy: 0.52552

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 141.00it/s]

valid Accuracy: 0.5332

------------------------------------------------------------

best validation accuracy is 53.3200 percent

***3***.

* How many trainable parameters are in the model? It is printed in the terminal.
  + There are 8069 parameters in this model
* What is the best training accuracy?
  + Best training Accuracy: 88.352 percent
* What is the best validation accuracy? Is it better than the best training accuracies?
  + Best validation accuracy is 85.3200 percent
  + No
* According to the training and validation accuracies, does the model overfit your training data?
  + As in this case, the validation accuracy is slightly lower than the training accuracy, it suggests that the model is not generalizing well. It may have overfit to the training data, meaning it has learned to memorize the training examples rather than capturing the underlying patterns..

There are 8069 parameters in this model

Use norm transformer for training

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:04<00:00, 45.55it/s]

train Accuracy: 0.62152

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 94.44it/s]

valid Accuracy: 0.762

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 78.25it/s]

train Accuracy: 0.77416

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 121.17it/s]

valid Accuracy: 0.8056

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 79.51it/s]

train Accuracy: 0.80576

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 121.38it/s]

valid Accuracy: 0.8096

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 79.44it/s]

train Accuracy: 0.8312

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 122.13it/s]

valid Accuracy: 0.8328

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 79.24it/s]

train Accuracy: 0.84656

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 122.04it/s]

valid Accuracy: 0.8372

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 75.71it/s]

train Accuracy: 0.85912

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 121.21it/s]

valid Accuracy: 0.85

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 76.06it/s]

train Accuracy: 0.8668

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 121.54it/s]

valid Accuracy: 0.852

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 78.11it/s]

train Accuracy: 0.87224

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 123.09it/s]

valid Accuracy: 0.8488

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 79.41it/s]

train Accuracy: 0.8788

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 122.49it/s]

valid Accuracy: 0.8532

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:02<00:00, 78.34it/s]

train Accuracy: 0.88352

100%|██████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 122.14it/s]

valid Accuracy: 0.8488

------------------------------------------------------------

best validation accuracy is 85.3200 percent

***4.***

* How many trainable parameters are in the model? It is printed in the terminal.
  + There are 29077 parameters in this model
* What is the best training accuracy?
  + Best training Accuracy: 89.48 percent
* What is the best validation accuracy? Is it better than the best training accuracies?
  + Best validation accuracy is 86.800 percent
  + Yes
* According to the training and validation accuracies, does the model overfit your training data?
  + It appears that the model may be overfitting the training data to some extent. The validation accuracy is slightly lower than the training accuracy, indicating that the model's performance on unseen data is not as good as its performance on the training data. it suggests that the model is not generalizing well.

There are 29077 parameters in this model

Use norm transformer for training

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.24it/s]

train Accuracy: 0.55112

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 52.66it/s]

valid Accuracy: 0.7312

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.41it/s]

train Accuracy: 0.74368

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.70it/s]

valid Accuracy: 0.7812

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 19.90it/s]

train Accuracy: 0.80016

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.84it/s]

valid Accuracy: 0.8116

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 20.47it/s]

train Accuracy: 0.81936

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.30it/s]

valid Accuracy: 0.8292

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 20.92it/s]

train Accuracy: 0.84208

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.03it/s]

valid Accuracy: 0.8108

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 21.42it/s]

train Accuracy: 0.8552

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.37it/s]

valid Accuracy: 0.8548

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 21.43it/s]

train Accuracy: 0.8704

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.88it/s]

valid Accuracy: 0.8636

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 21.30it/s]

train Accuracy: 0.87816

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.31it/s]

valid Accuracy: 0.868

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 20.42it/s]

train Accuracy: 0.8884

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 51.93it/s]

valid Accuracy: 0.8672

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:09<00:00, 21.29it/s]

train Accuracy: 0.8948

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 59.97it/s]

valid Accuracy: 0.8676

------------------------------------------------------------

best validation accuracy is 86.8000 percent

***5.***

* How many trainable parameters are in the model? It is printed in the terminal.
  + There are 29077 parameters in this model
* What is the best training accuracy?
  + Best training Accuracy: 86.86864 percent
* What is the best validation accuracy? Is it better than the best training accuracies?
  + Best validation accuracy is 86.5200 percent
  + No
* According to the training and validation accuracies, does the model overfit your training data?
  + Since the validation accuracy is slightly lower than the training accuracy, it suggests that the model may be overfitting the training data to some extent. Overfitting occurs when a model becomes too specialized in learning the details and noise in the training data, making it less effective at generalizing to new, unseen data. In this case, the small difference between the best training and validation accuracies indicates a relatively small degree of overfitting.

There are 29077 parameters in this model

Use aug transformer for training

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 18.33it/s]

train Accuracy: 0.54464

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.01it/s]

valid Accuracy: 0.7356

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 18.47it/s]

train Accuracy: 0.73456

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.02it/s]

valid Accuracy: 0.782

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 18.85it/s]

train Accuracy: 0.7844

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.10it/s]

valid Accuracy: 0.8152

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.01it/s]

train Accuracy: 0.81224

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.04it/s]

valid Accuracy: 0.8276

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.06it/s]

train Accuracy: 0.82416

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.80it/s]

valid Accuracy: 0.8504

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 18.95it/s]

train Accuracy: 0.84048

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.19it/s]

valid Accuracy: 0.8508

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.55it/s]

train Accuracy: 0.85128

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.02it/s]

valid Accuracy: 0.8576

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.57it/s]

train Accuracy: 0.85768

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 56.32it/s]

valid Accuracy: 0.8676

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.19it/s]

train Accuracy: 0.86416

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.50it/s]

valid Accuracy: 0.8744

100%|█████████████████████████████████████████████████████████████████| 196/196 [00:10<00:00, 19.00it/s]

train Accuracy: 0.86864

100%|███████████████████████████████████████████████████████████████████| 40/40 [00:00<00:00, 55.23it/s]

valid Accuracy: 0.8652

------------------------------------------------------------

best validation accuracy is 87.4400 percent