Only Finetuning

Bert Base Uncased is a transformers model pretrained on a large corpus of English language using a masked language modeling (MLM) objective in a self-supervised fashion. This means it was pretrained on the raw texts only, with no humans labelling them in any way with an automatic process to generate inputs and labels from those texts. This model is uncased: it does not make a difference between english and English. There is a cased model version as well and I use both the versions to compare the results.

Masked language modeling (MLM) takes a sentence, the model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words. It allows the model to learn a bidirectional representation of the sentence.

FineTuning of the Bert Model
Bert-base-uncased model is fine tuned on the training data set (df_train) and parameters are finetuned using the validation dataset. Based on the best model selected as per validation dataset (df_val), test accuracy is measured.

Tokeniser and pretrained model are loaded -

```
tokenizer = BertTokenizer.from_pretrained(PRE_TRAINED_MODEL_NAME)

bert_model = BertModel.from_pretrained(PRE_TRAINED_MODEL_NAME, return_dict=False)
```

Data loaders are specified for both training and val datasets. Tokenizers and other parameters are provided.

```
train_data_loader = create_data_loader(df_train, tokenizer, MAX_LEN, BATCH_SIZE)

val_data_loader = create_data_loader(df_val, tokenizer, MAX_LEN, BATCH_SIZE)
```

Dataset Classes – {0.0, 0.2, 0.4, 0.6, 0.8, 1.0}
There are very few examples with values other than the 6 values (98 examples in training data set of 17500). Number of classes used is 6.

Training for the whole batch	**Evaluating** for the dataset in the batch
1. Load batch of training data set 2. Forward pass 3. Compute Loss and accuracy 4. Backward pass 5. Optimization	1. Generate output 2. Predict using the SoftMax layer 3. Compute loss and accuracy

Evaluation after each Epoch helps in identifying the Best model that is stored for later use.
Model can be used and loaded using PyTorch library function save and load.

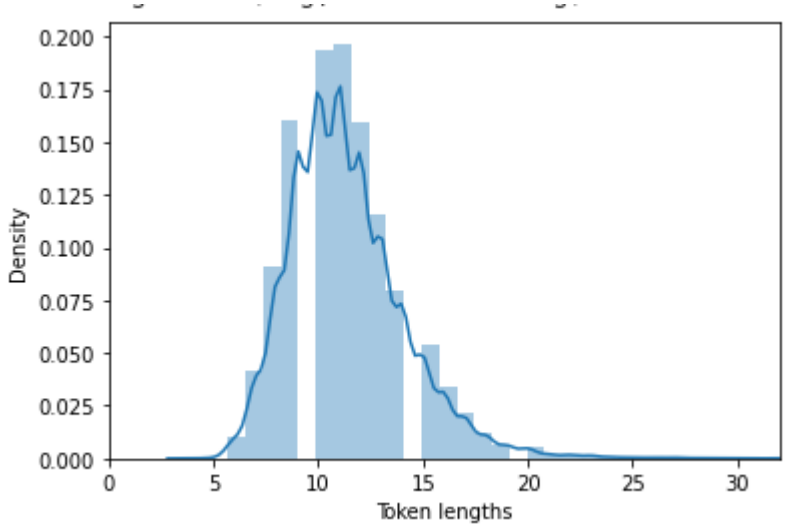
Test Accuracy is computed on the best model using the same evaluation function.
IPYNB file is attached along with the report.

Hyper Parameter Tuning - Selection of best model

Multiple experiments were performed on different parameters of the Bert Training. These parameters are

1. Max_length – Length of the token vector is dependent on the input textual data. As is visible from the chart below, length of 32 was considered sufficient and I experimented with lower Max_length was well. However, the Validation accuracy worsened for these scenarios. Max_length was retained as 32 in remaining experiments.

Length of Queries – Distribution as guide to deciding Token Length



Max_Length	Best Validation Accuracy	Epoch in which reached
32 (Experiment 1)	80.51%	1
24 (Experiment 2)	79.71%	1
20 (Experiment 3)	79.71%	1

Other parameters were kept fixed during this experiment - Number of Epochs = 5
Pretrained Model = bert-base-uncased, Batch size = 16, num_warmup_steps=0 and lr=2e-5

2. Choice of Bert Base Cased or Uncased – I started with initially the Bert-Base-Uncased as the initial pretrained model and replaced it with Bert Base Cased to see any potential impact on validation accuracy. Validation accuracy dipped to 80.08% from 80.51%.

Pretrained Model	Best Validation Accuracy	Epoch in which reached
Bert-Base-Uncased (Experiment 1)	80.51%	1
Bert-Base-Uncased (Experiment 4)	80.08%	2

Other parameters were kept fixed during this experiment - Number of Epochs = 5, Max_length = 32, Batch size = 16, num_warmup_steps=0 and lr=2e-5

3. Learning rate – Learning rate base case was 2e-5, multiple experiments were conducted on different learning rates, however, the performance was not better.

Learning Rate	Best Validation Accuracy	Epoch in which reached

2e-5 (Experiment 1)	80.51%	1
2e-6 (Experiment 5)	79.39%	1
2e-7 (Experiment 6)	79.69%	1
4e-5 (Experiment 7)	79.23%	2
2e-4 (Experiment 8)	76.69%	1

Other parameters were kept fixed during this experiment - Number of Epochs = 5, Max_length = 32, Pretrained Model = bert-base-uncased, **Batch size = 16,** num_warmup_steps=0

Batch size change to 20 resulted in higher validation accuracy. Further experiments were done on learning rate to see if there was any impact. No improvement was observed.

Learning Rate	Best Validation Accuracy	Epoch in which reached
2e-5 (Experiment 10)	81.33%	1
0.5 *1e-5 (Experiment 12)	81.04%	2
4e-5 (Experiment 13)	76.69%	1

Other parameters were kept fixed during this experiment - Number of Epochs = 5, Max_length = 32, Pretrained Model = bert-base-uncased, **Batch size = 20**, num_warmup_steps=0

4. Batch size – The number of training samples which are given to GPU in a single batch is the batch size. I started with batch size of 16. However, performance of validation accuracy showed lot of variations with batch size (relative to other parameters). Multiple experiment was conducted on this parameter. Validation accuracy increased to 81.33% when batch size was increased to 24 and further increased to 81.60% on batch size of 32. Further increases in batch size did not improve the validation accuracy.

Batch size	Best Validation Accuracy	Epoch in which reached
16 (Experiment 1)	**80.51%**	1
12 (Experiment 9)	79.44%	1
20 (Experiment 10)	**81.33%**	1
24 (Experiment 11)	79.81%	1
32 (Experiment 17)	**81.60%**	1
48 (Experiment 18)	80.99%	1
56 (Experiment 19)	81.09%	1
64 (Experiment 20)	81.17%	1

Other parameters were kept fixed during this experiment - Number of Epochs = 5, Max_length = 32, Pretrained Model = bert-base-uncased, num_warmup_steps=0, learning rate = 2e-5

5. Number of warm up steps - The number of warm up steps for which the training is done. We started with 0 warm up steps and then compared multiple parameter values. No improvement in the validation accuracy was seen.

Number of warm up steps	Best Validation Accuracy	Epoch in which reached

0 (Experiment 10)	81.33%	1
100 (Experiment 14)	80.80%	1
200 (Experiment 15)	80.29%	5
50 (Experiment 16)	80.61%	1

Other parameters were kept fixed during this experiment - Number of Epochs = 5, Max_length = 32, Pretrained Model = bert-base-uncased, Batch size = 20, learning rate = 2e-5

6. Number of training Epochs – Training for 5 Epochs was seen was sufficient as training accuracy by Epoch 5 was crossing 95% and validation accuracy was either flat or showed a declining trend with more epochs of training. Further experiments were done on number of Epochs to find a better parameter value.

Number of Epochs	Best Validation Accuracy	Epoch in which reached
5 (Experiment 17)	**81.60%**	1
4 (Experiment 21)	81.41%	1
6 (Experiment 22)	81.31%	1
8 (Experiment 23)	80.85%	1
2(Experiment 24)	**81.71%**	1

Other parameters were kept fixed during this experiment - Max_length = 32, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, learning rate = 2e-5

Best Model Selection and Test Accuracy on Best Model
The best model based on highest validation accuracy of 81.71% had the following parameters

Max_length	32
Pretrained Model	bert-base-uncased
Batch size	32
num_warmup_steps	0
learning rate	2e-5
Number of Epochs	2 (best model saved after Epoch 1)

Final Test Accuracy Obtained on this model is 80.7% (validation accuracy 81.71%)

As compared to other approaches known for this dataset, accuracy of BERT transformed based model is much better. Detailed hyperparameter tuning results are there in appendix.

Pretraining + Finetuning

Retraining an already pretrained model essentially means providing the training set and let the BERT relearn the language model using the training data set. Since training set often differs from

the Book corpus on which BERT was originally trained, it enables the model to adjust the language model accordingly.

Masked language modeling (MLM) takes a sentence, the model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words. It allows the model to learn a bidirectional representation of the sentence.

Bert-base-uncased model is first retrained on the training data set (df_train). Labels are removed, text converted to lower case. BERT Model loaded using BertForMaskedLM architecture.
Two variants of Retraining are used –

1. Using the whole training dataset,
2. Using those queries for which the score is more than 0.2 (10929 queries out of 17500 queries qualified). Intent was to use only the good queries to train the language model.
3. Using those queries for which the score is more than 0.6 (6821 queries out of 17500 queries qualified). Intent was to use only the best queries to train the language model.

Tokeniser and pretrained model are loaded -

```
PRE_TRAINED_MODEL_NAME = 'bert-base-uncased'

tokenizer = BertTokenizer.from_pretrained(PRE_TRAINED_MODEL_NAME)

bert_model = BertForMaskedLM.from_pretrained(PRE_TRAINED_MODEL_NAME, return_dict=False)
```

Data is loaded using a predefined class LineByLineTextDataset

```
from transformers import LineByLineTextDataset

dataset = LineByLineTextDataset(
tokenizer=tokenizer,
file_path="./train_text_goodscores.tsv",
block_size=32
)
```

DataCollatorForLanguageModeling class is used for creating the dataloader

```
from transformers import DataCollatorForLanguageModeling

data_collator = DataCollatorForLanguageModeling(
tokenizer=tokenizer, mlm=True, mlm_probability=0.15
)
```

Training arguments and Trainer details are specified.

```
from transformers import Trainer, TrainingArguments

training_args = TrainingArguments(
output_dir="./bert_base_uncased",
```

```
overwrite_output_dir=True,
num_train_epochs=5,
per_device_train_batch_size=32,
save_steps=500,
save_total_limit=2,
seed=1
)

trainer = Trainer(
model=bert_model,
args=training_args,
data_collator=data_collator,
train_dataset=dataset
)
```

After the training is complete, model is saved and is then finetuned using the process followed earlier.

In the fine tuning, evaluation after each Epoch helps in identifying the Best model that is stored for later use. Model can be saved and loaded using PyTorch library function save and load.

Test Accuracy is computed on the best model using the same evaluation function.
IPYNB file is attached along with the report.

Accuracy on Validation Dataset for different parameters of **Retraining** -

	Batch size = 32, 5 Epochs	Batch size = 32, 7 Epochs	Batch size = 32, 2 Epochs	Batch size =16, 7 Epochs	Batch size = 48, 7 Epochs
Training set with labels more than 0.2	81.25% (Experiment 27)	81.92% (Experiment 25)	81.28% (Experiment 26)	81.09% (Experiment 28)	81.68% (Experiment 29)
Full Training set	81.17% (Experiment 30)	80.75% (Experiment 31)	81.07% (Experiment 32)		
Training set with labels more than 0.6	**82.05%** (Experiment 33)	82.00% (Experiment 34)	81.20% (Experiment 35)		

Parameters used for finetuning are same across all the Retraining attempts, MAX_LEN = 32, BATCH_SIZE = 32, EPOCHS = 2, mylearn_rate = 2e-5, NUM_CLASSES = 6

From the above table, it is clear that Training Set when poor quality examples are removed, performs better than the total training set. And the best training examples (labels > 0.6) performs the best, with validation accuracy of 82.05% for 5 Epochs and 32 batch size.

Finetuning of Model – Hyper Parameter Tuning
Based on the learning that batch size and number of epochs have higher influence in terms of impact, three value of batch sizes during finetuning and Epochs are tried.

The parameters during the finetuning stage are modified to validate if there is an improvement possible -

	Batch size = 32, 2 Epochs	Batch size = 32, 5 Epochs	Batch size = 32, 7 Epochs	Batch size =16, 2 Epochs	Batch size = 48, 2 Epochs
Training set with labels more than 0.6	**82.05%** (Experiment 33)	80.59% (Experiment 36)	81.33% (Experiment 37)	80.13% (Experiment 38)	80.96% (Experiment 39)

Hyper parameter tuning of parameters did not improve the accuracy on validation data.

For the final set of parameters, test accuracy was computed which was 81.30%, validation accuracy of 82.05%.

Both the validation and the test accuracy are better after retraining of the BERT model with the training set hence validating the hypothesis. Detailed hyperparameter tuning results are there in appendix.

Appendix
Experiment 1

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 16
num_warmup_steps=0
lr=2e-5

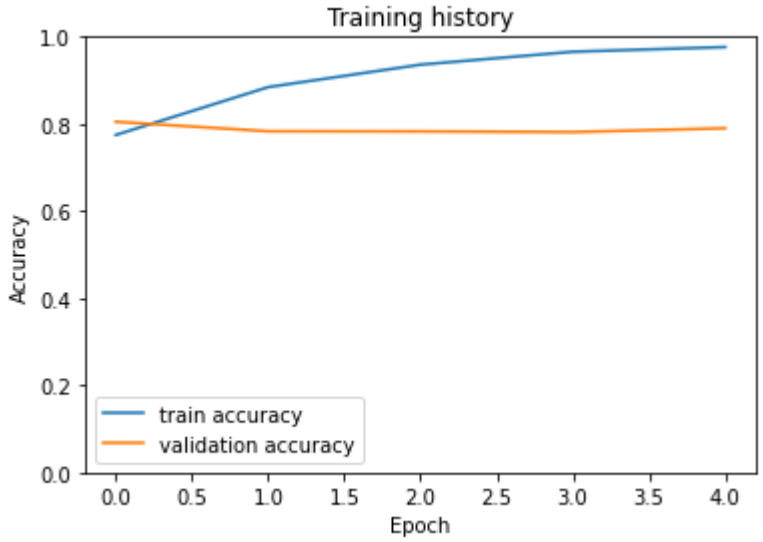
Epoch: 01
Train Loss: 0.449 | Train Acc: 77.46%
Val. Loss: 0.385 | Val. Acc: 80.51%

Epoch: 02
Train Loss: 0.279 | Train Acc: 88.45%
Val. Loss: 0.436 | Val. Acc: 78.32%

Epoch: 03
Train Loss: 0.191 | Train Acc: 93.61%
Val. Loss: 0.586 | Val. Acc: 78.29%

Epoch: 04
Train Loss: 0.125 | Train Acc: 96.58%
Val. Loss: 0.779 | Val. Acc: 78.13%

Epoch: 05
Train Loss: 0.095 | Train Acc: 97.65%
Val. Loss: 1.000 | Val. Acc: 79.01%


Training history
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
train accuracy
validation accuracy
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Epoch

Experiment 2

Length of Queries = **24**
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size (for both train and val) = 16
num_warmup_steps=0
lr=2e-5

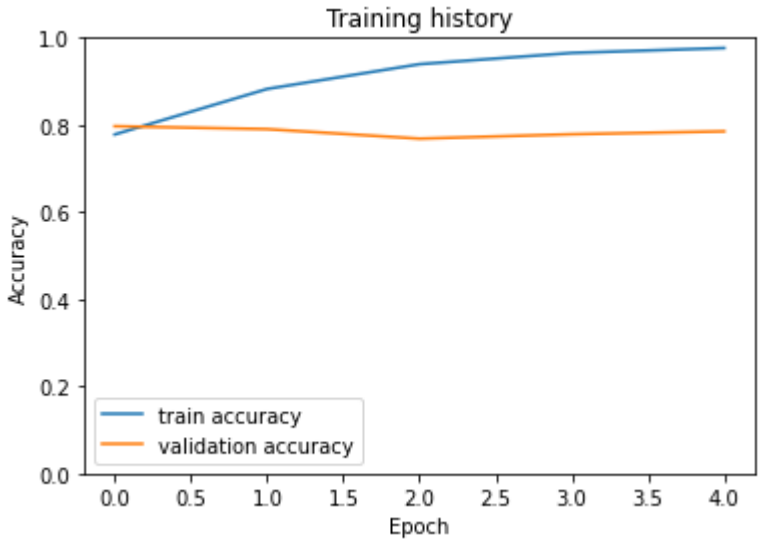
Epoch: 01
Train Loss: 0.441 | Train Acc: 77.82%
Val. Loss: 0.385 | Val. Acc: 79.71%

Epoch: 02
Train Loss: 0.279 | Train Acc: 88.26%
Val. Loss: 0.461 | Val. Acc: 79.04%

Epoch: 03
Train Loss: 0.184 | Train Acc: 93.95%
Val. Loss: 0.646 | Val. Acc: 76.85%

Epoch: 04
Train Loss: 0.123 | Train Acc: 96.53%
Val. Loss: 0.880 | Val. Acc: 77.87%

Epoch: 05
Train Loss: 0.094 | Train Acc: 97.65%
Val. Loss: 1.058 | Val. Acc: 78.53%



Experiment 3

Length of Queries = **20**
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size (for both train and val) = 16
num_warmup_steps=0
lr=2e-5

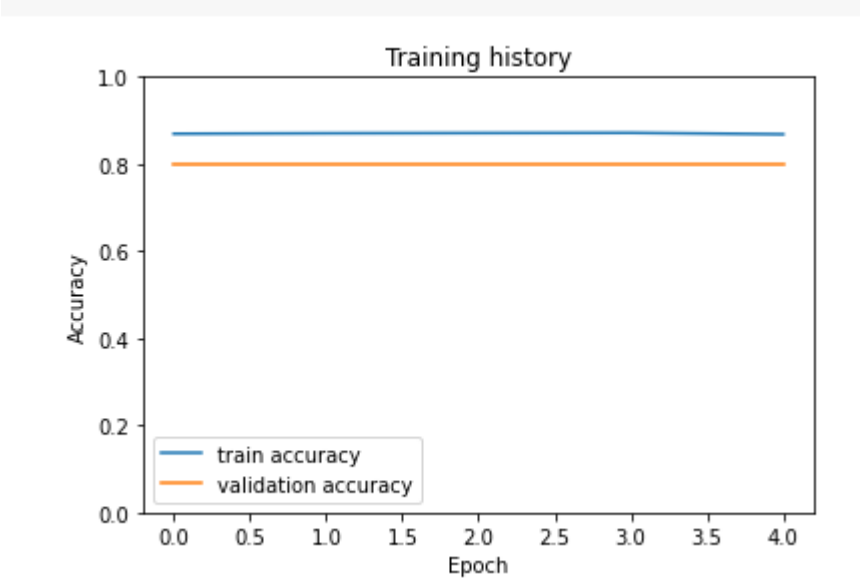
Epoch: 01
Train Loss: 0.292 | Train Acc: 86.93%
Val. Loss: 0.385 | Val. Acc: 79.71%

Epoch: 02
Train Loss: 0.294 | Train Acc: 87.05%
Val. Loss: 0.385 | Val. Acc: 79.71%

Epoch: 03
Train Loss: 0.293 | Train Acc: 87.10%
Val. Loss: 0.385 | Val. Acc: 79.71%

Epoch: 04
Train Loss: 0.292 | Train Acc: 87.15%
Val. Loss: 0.385 | Val. Acc: 79.71%

Epoch: 05
Train Loss: 0.295 | Train Acc: 86.82%
Val. Loss: 0.385 | Val. Acc: 79.71%



Experiment 4

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = **bert-base-cased**
Batch size = 16
num_warmup_steps=0
lr=2e-5

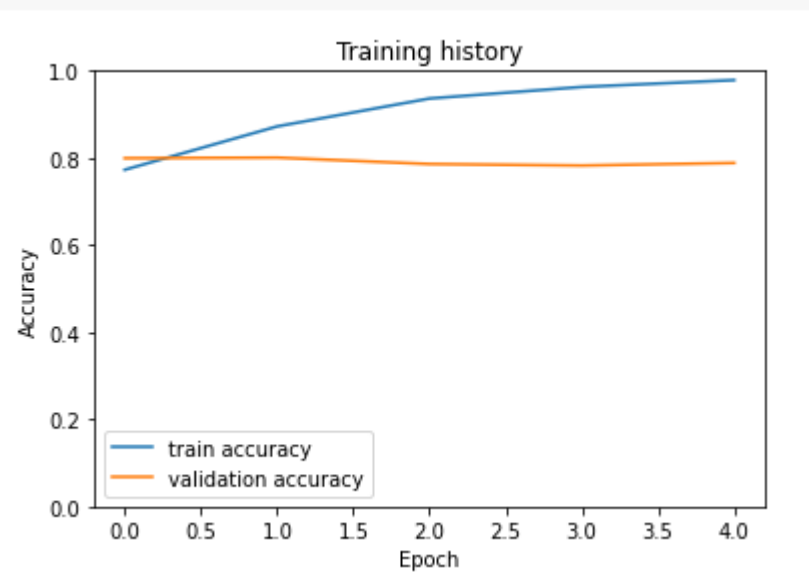
Epoch: 01
Train Loss: 0.462 | Train Acc: 77.29%
Val. Loss: 0.395 | Val. Acc: 79.95%

Epoch: 02
Train Loss: 0.297 | Train Acc: 87.25%
Val. Loss: 0.479 | Val. Acc: 80.08%

Epoch: 03 | Epoch Time: 1m 35s
Train Loss: 0.187 | Train Acc: 93.65%
Val. Loss: 0.710 | Val. Acc: 78.61%

Epoch: 04 | Epoch Time: 1m 34s
Train Loss: 0.127 | Train Acc: 96.30%
Val. Loss: 0.945 | Val. Acc: 78.27%

Epoch: 05 | Epoch Time: 1m 34s
Train Loss: 0.086 | Train Acc: 97.87%
Val. Loss: 1.148 | Val. Acc: 78.85%



Experiment 5

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 16
num_warmup_steps=0
lr=2e-6

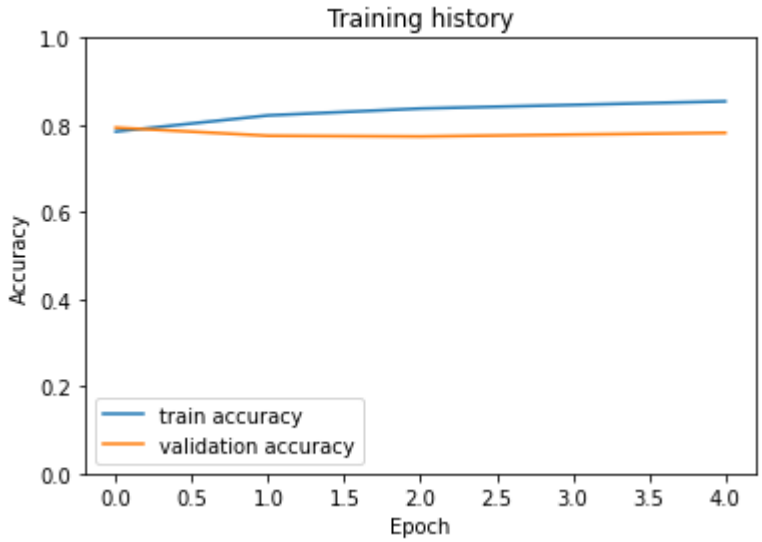
Epoch: 01
Train Loss: 0.441 | Train Acc: 78.49%
Val. Loss: 0.393 | Val. Acc: **79.39%**

Epoch: 02
Train Loss: 0.372 | Train Acc: 82.19%
Val. Loss: 0.416 | Val. Acc: 77.52%

Epoch: 03
Train Loss: 0.348 | Train Acc: 83.77%
Val. Loss: 0.432 | Val. Acc: 77.36%

Epoch: 04
Train Loss: 0.331 | Train Acc: 84.60%
Val. Loss: 0.434 | Val. Acc: 77.79%

Epoch: 05
Train Loss: 0.323 | Train Acc: 85.42%
Val. Loss: 0.434 | Val. Acc: 78.19%



Experiment 6

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 16
num_warmup_steps=0
lr=2e-7

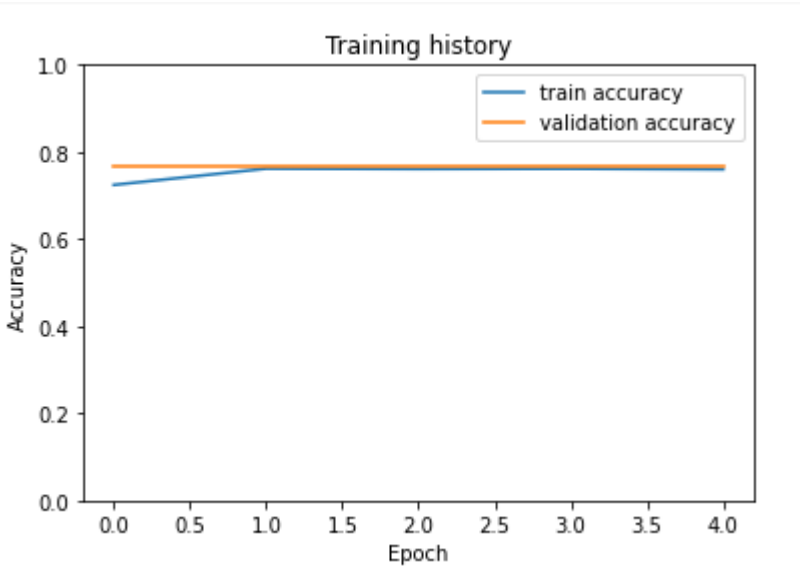
Epoch: 01 | Epoch Time: 1m 35s
Train Loss: 1.007 | Train Acc: 72.43%
Val. Loss: 0.731 | Val. Acc: **76.69%**

Epoch: 02 | Epoch Time: 1m 35s
Train Loss: 0.726 | Train Acc: 76.18%
Val. Loss: 0.625 | Val. Acc: 76.69%

Epoch: 03 | Epoch Time: 1m 35s
Train Loss: 0.658 | Train Acc: 76.10%
Val. Loss: 0.576 | Val. Acc: 76.69%

Epoch: 04 | Epoch Time: 1m 35s
Train Loss: 0.625 | Train Acc: 76.15%
Val. Loss: 0.556 | Val. Acc: 76.69%

Epoch: 05 | Epoch Time: 1m 35s
Train Loss: 0.610 | Train Acc: 76.01%
Val. Loss: 0.550 | Val. Acc: 76.69%



Experiment 7

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 16
num_warmup_steps=0
lr=4e-6

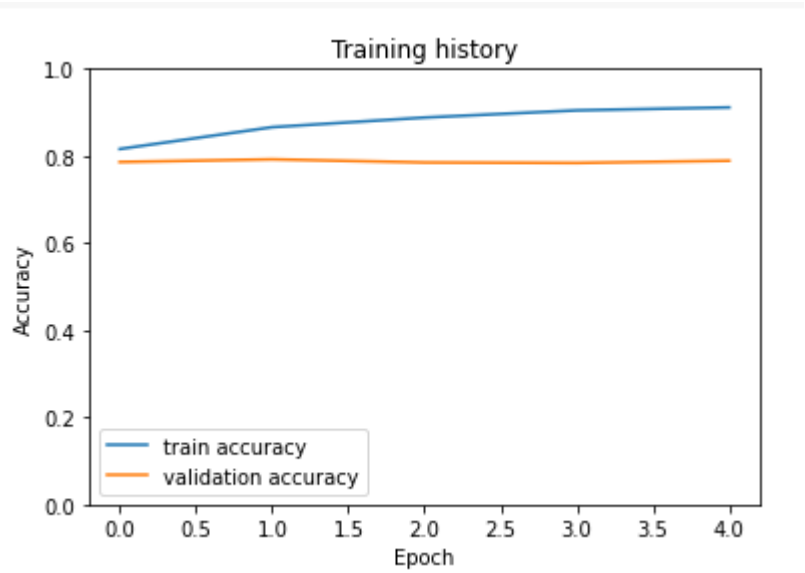
Epoch: 01 | Epoch Time: 1m 35s
Train Loss: 0.378 | Train Acc: 81.59%
Val. Loss: 0.386 | Val. Acc: 78.61%

Epoch: 02 | Epoch Time: 1m 35s
Train Loss: 0.302 | Train Acc: 86.61%
Val. Loss: 0.403 | Val. Acc: **79.23%**

Epoch: 03 | Epoch Time: 1m 35s
Train Loss: 0.264 | Train Acc: 88.82%
Val. Loss: 0.463 | Val. Acc: 78.51%

Epoch: 04 | Epoch Time: 1m 35s
Train Loss: 0.240 | Train Acc: 90.47%
Val. Loss: 0.526 | Val. Acc: 78.40%

Epoch: 05 | Epoch Time: 1m 35s
Train Loss: 0.233 | Train Acc: 91.14%
Val. Loss: 0.580 | Val. Acc: 78.91%



Experiment 8

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 16
num_warmup_steps=0
lr=2e-4

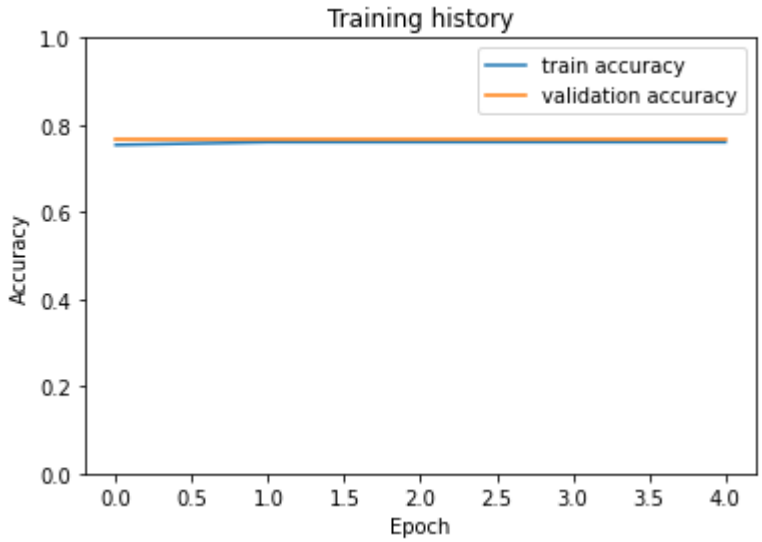
Epoch: 01
Train Loss: 0.578 | Train Acc: 75.42%
Val. Loss: 0.545 | Val. Acc: **76.69%**

Epoch: 02
Train Loss: 0.556 | Train Acc: 76.06%
Val. Loss: 0.547 | Val. Acc: 76.69%

Epoch: 03 | Epoch Time: 1m 35s
Train Loss: 0.554 | Train Acc: 76.06%
Val. Loss: 0.546 | Val. Acc: 76.69%

Epoch: 04 | Epoch Time: 1m 36s
Train Loss: 0.553 | Train Acc: 76.06%
Val. Loss: 0.545 | Val. Acc: 76.69%

Epoch: 05 | Epoch Time: 1m 35s
Train Loss: 0.553 | Train Acc: 76.06%
Val. Loss: 0.544 | Val. Acc: 76.69%



Experiment 9

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 12
num_warmup_steps=0
lr=2e-5

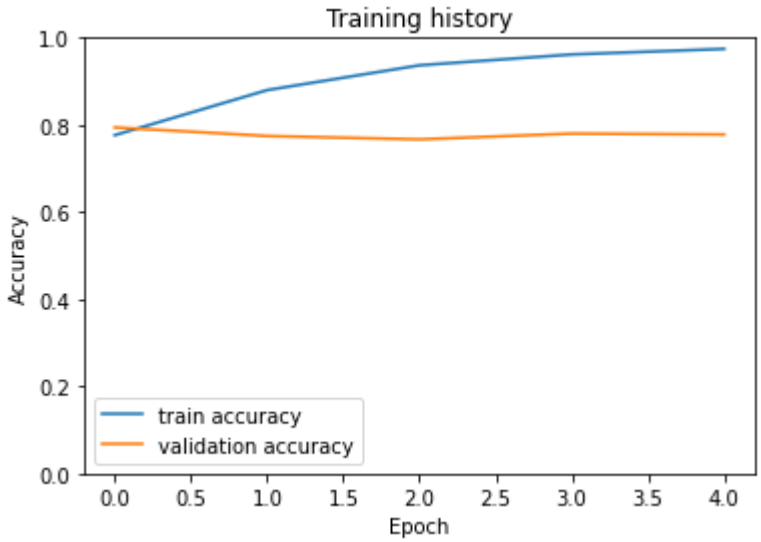
Epoch: 01 | Epoch Time: 1m 55s
Train Loss: 0.451 | Train Acc: 77.65%
Val. Loss: 0.391 | Val. Acc: **79.44%**

Epoch: 02 | Epoch Time: 1m 54s
Train Loss: 0.293 | Train Acc: 87.99%
Val. Loss: 0.485 | Val. Acc: 77.47%

Epoch: 03 | Epoch Time: 1m 55s
Train Loss: 0.198 | Train Acc: 93.69%
Val. Loss: 0.735 | Val. Acc: 76.69%

Epoch: 04 | Epoch Time: 1m 54s
Train Loss: 0.137 | Train Acc: 96.18%
Val. Loss: 0.955 | Val. Acc: 78.03%

Epoch: 05 | Epoch Time: 1m 53s
Train Loss: 0.103 | Train Acc: 97.45%
Val. Loss: 1.049 | Val. Acc: 77.81%



Experiment 10

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 20
num_warmup_steps=0
lr=2e-5

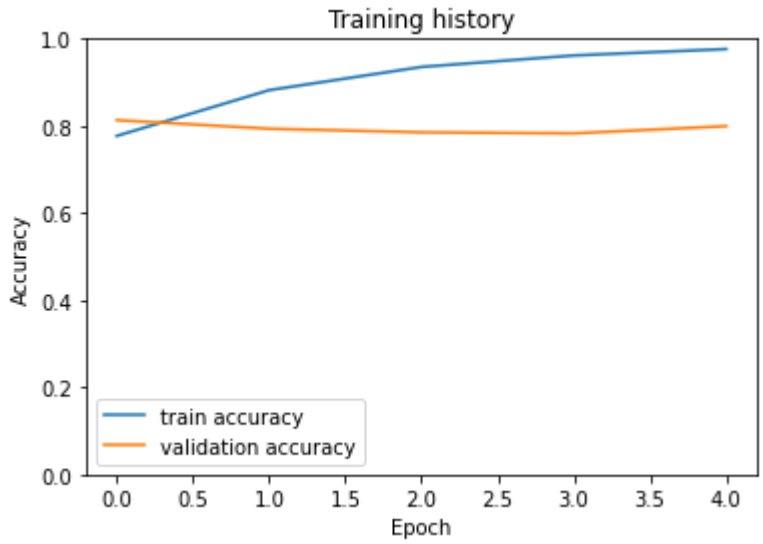
Epoch: 01 | Epoch Time: 1m 30s
Train Loss: 0.443 | Train Acc: 77.71%
Val. Loss: 0.373 | Val. Acc: 81.33%

Epoch: 02 | Epoch Time: 1m 30s
Train Loss: 0.274 | Train Acc: 88.21%
Val. Loss: 0.445 | Val. Acc: 79.36%

Epoch: 03 | Epoch Time: 1m 30s
Train Loss: 0.181 | Train Acc: 93.54%
Val. Loss: 0.557 | Val. Acc: 78.53%

Epoch: 04 | Epoch Time: 1m 30s
Train Loss: 0.130 | Train Acc: 96.18%
Val. Loss: 0.782 | Val. Acc: 78.29%

Epoch: 05 | Epoch Time: 1m 30s
Train Loss: 0.090 | Train Acc: 97.65%
Val. Loss: 0.902 | Val. Acc: 79.95%



Experiment 11

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 24
num_warmup_steps=0
lr=2e-5

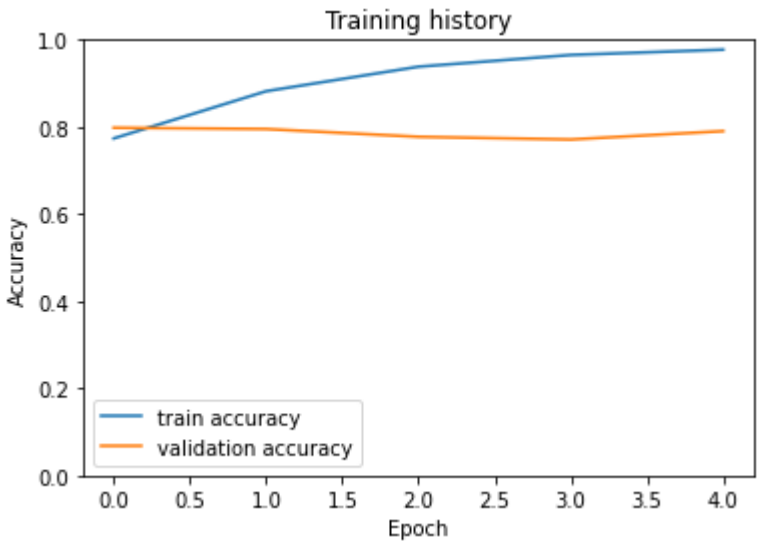
Epoch: 01 | Epoch Time: 1m 24s
Train Loss: 0.439 | Train Acc: 77.37%
Val. Loss: 0.377 | Val. Acc: **79.81%**

Epoch: 02 | Epoch Time: 1m 24s
Train Loss: 0.267 | Train Acc: 88.20%
Val. Loss: 0.436 | Val. Acc: 79.52%

Epoch: 03 | Epoch Time: 1m 24s
Train Loss: 0.166 | Train Acc: 93.83%
Val. Loss: 0.578 | Val. Acc: 77.73%

Epoch: 04 | Epoch Time: 1m 24s
Train Loss: 0.118 | Train Acc: 96.52%
Val. Loss: 0.784 | Val. Acc: 77.12%

Epoch: 05 | Epoch Time: 1m 24s
Train Loss: 0.081 | Train Acc: 97.74%
Val. Loss: 0.990 | Val. Acc: 79.04%



Experiment 12

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 20
num_warmup_steps=0
lr=0.5 * 1e-5

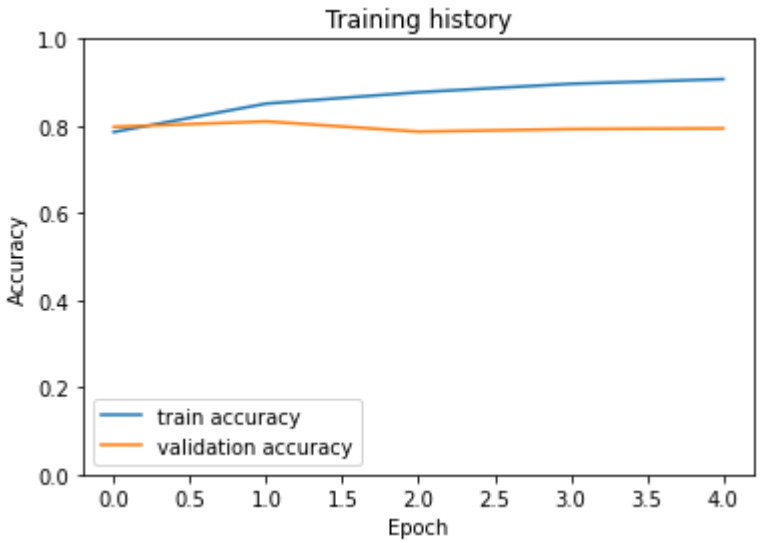
Epoch: 01 | Epoch Time: 1m 31s
Train Loss: 0.437 | Train Acc: 78.60%
Val. Loss: 0.376 | Val. Acc: 79.79%

Epoch: 02 | Epoch Time: 1m 30s
Train Loss: 0.323 | Train Acc: 85.13%
Val. Loss: 0.373 | Val. Acc: 81.04%

Epoch: 03 | Epoch Time: 1m 31s
Train Loss: 0.278 | Train Acc: 87.74%
Val. Loss: 0.424 | Val. Acc: 78.69%

Epoch: 04 | Epoch Time: 1m 31s
Train Loss: 0.252 | Train Acc: 89.67%
Val. Loss: 0.463 | Val. Acc: 79.28%

Epoch: 05 | Epoch Time: 1m 30s
Train Loss: 0.231 | Train Acc: 90.74%
Val. Loss: 0.528 | Val. Acc: 79.44%



Experiment 13

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 20
num_warmup_steps=0
lr= 4e-5

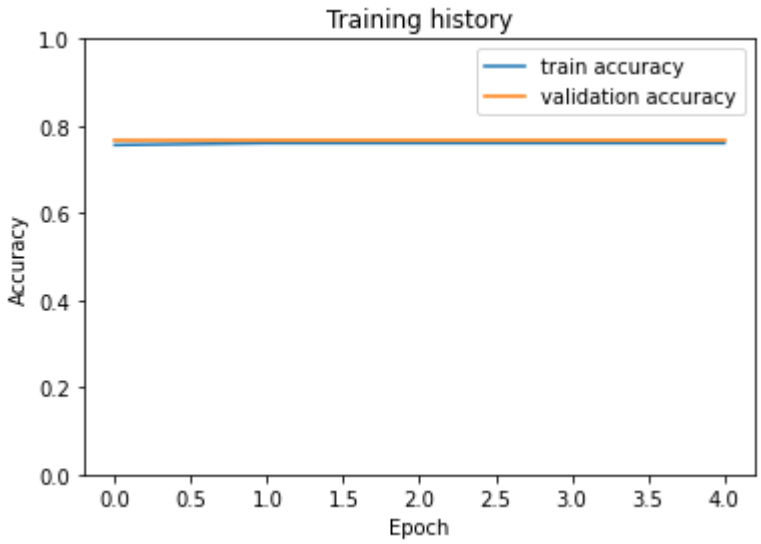
Epoch: 01 | Epoch Time: 1m 31s
Train Loss: 0.568 | Train Acc: 75.63%
Val. Loss: 0.549 | Val. Acc: **76.69%**

Epoch: 02 | Epoch Time: 1m 32s
Train Loss: 0.555 | Train Acc: 76.06%
Val. Loss: 0.544 | Val. Acc: 76.69%

Epoch: 03 | Epoch Time: 1m 31s
Train Loss: 0.555 | Train Acc: 76.06%
Val. Loss: 0.544 | Val. Acc: 76.69%

Epoch: 04 | Epoch Time: 1m 31s
Train Loss: 0.553 | Train Acc: 76.06%
Val. Loss: 0.544 | Val. Acc: 76.69%

Epoch: 05 | Epoch Time: 1m 32s
Train Loss: 0.553 | Train Acc: 76.06%
Val. Loss: 0.544 | Val. Acc: 76.69%



Experiment 14

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 20
num_warmup_steps=100
lr=2e-5

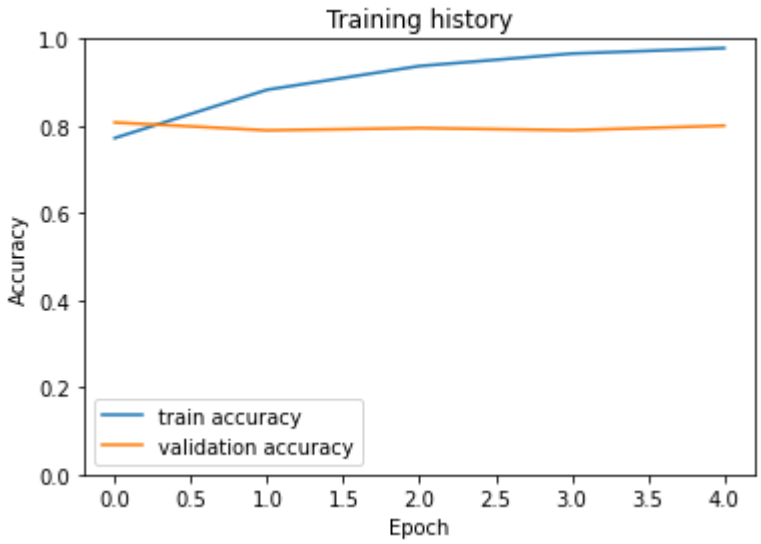
Epoch: 01 | Epoch Time: 1m 32s
Train Loss: 0.462 | Train Acc: 77.22%
Val. Loss: 0.374 | Val. Acc: **80.80%**

Epoch: 02 | Epoch Time: 1m 31s
Train Loss: 0.267 | Train Acc: 88.31%
Val. Loss: 0.461 | Val. Acc: 78.99%

Epoch: 03 | Epoch Time: 1m 31s
Train Loss: 0.171 | Train Acc: 93.75%
Val. Loss: 0.572 | Val. Acc: 79.52%

Epoch: 04 | Epoch Time: 1m 31s
Train Loss: 0.117 | Train Acc: 96.61%
Val. Loss: 0.812 | Val. Acc: 79.01%

Epoch: 05 | Epoch Time: 1m 31s
Train Loss: 0.080 | Train Acc: 97.86%
Val. Loss: 0.994 | Val. Acc: 80.05%



Experiment 15

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 20
num_warmup_steps=200
lr=2e-5

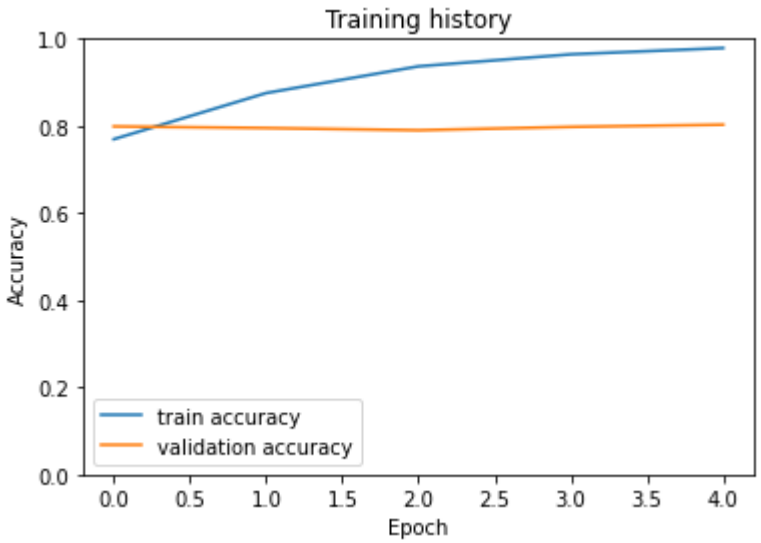
Epoch: 01 | Epoch Time: 1m 32s
Train Loss: 0.477 | Train Acc: 76.95%
Val. Loss: 0.379 | Val. Acc: 79.89%

Epoch: 02 | Epoch Time: 1m 31s
Train Loss: 0.276 | Train Acc: 87.55%
Val. Loss: 0.449 | Val. Acc: 79.52%

Epoch: 03 | Epoch Time: 1m 31s
Train Loss: 0.170 | Train Acc: 93.68%
Val. Loss: 0.557 | Val. Acc: 79.01%

Epoch: 04 | Epoch Time: 1m 32s
Train Loss: 0.117 | Train Acc: 96.44%
Val. Loss: 0.741 | Val. Acc: 79.79%

Epoch: 05 | Epoch Time: 1m 31s
Train Loss: 0.081 | Train Acc: 97.88%
Val. Loss: 0.974 | Val. Acc: **80.29%**



Experiment 16

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-cased
Batch size = 20
num_warmup_steps=50
lr=2e-5

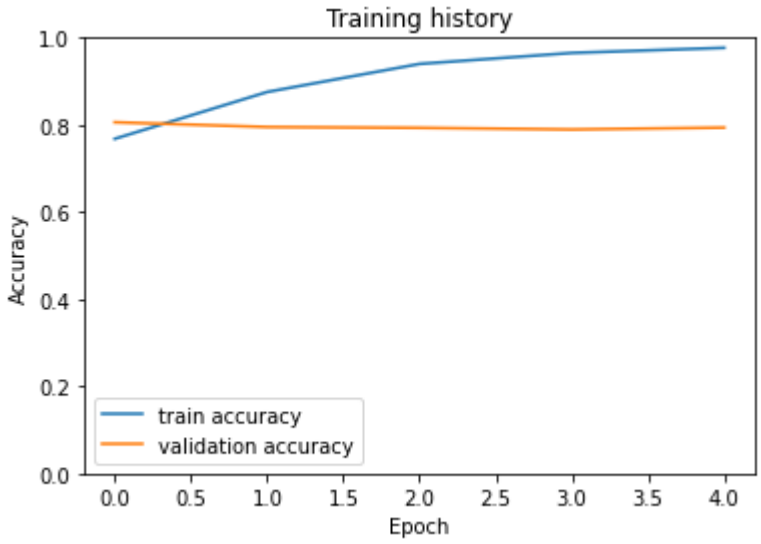
Epoch: 01 | Epoch Time: 1m 30s
Train Loss: 0.476 | Train Acc: 76.81%
Val. Loss: 0.383 | Val. Acc: **80.61%**

Epoch: 02 | Epoch Time: 1m 29s
Train Loss: 0.283 | Train Acc: 87.56%
Val. Loss: 0.460 | Val. Acc: 79.52%

Epoch: 03 | Epoch Time: 1m 30s
Train Loss: 0.170 | Train Acc: 94.01%
Val. Loss: 0.601 | Val. Acc: 79.33%

Epoch: 04 | Epoch Time: 1m 30s
Train Loss: 0.114 | Train Acc: 96.53%
Val. Loss: 0.790 | Val. Acc: 78.96%

Epoch: 05 | Epoch Time: 1m 30s
Train Loss: 0.082 | Train Acc: 97.70%
Val. Loss: 1.030 | Val. Acc: 79.39%



Experiment 17

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 32
num_warmup_steps=0
lr=2e-5

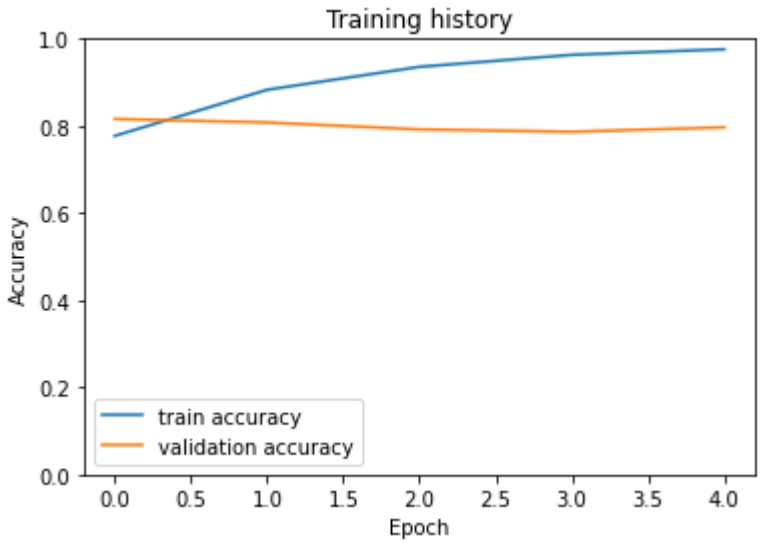
Epoch: 01 | Epoch Time: 1m 17s
Train Loss: 0.437 | Train Acc: 77.71%
Val. Loss: 0.371 | Val. Acc: 81.60%

Epoch: 02 | Epoch Time: 1m 17s
Train Loss: 0.270 | Train Acc: 88.30%
Val. Loss: 0.419 | Val. Acc: 80.80%

Epoch: 03 | Epoch Time: 1m 17s
Train Loss: 0.171 | Train Acc: 93.57%
Val. Loss: 0.500 | Val. Acc: 79.20%

Epoch: 04 | Epoch Time: 1m 17s
Train Loss: 0.115 | Train Acc: 96.32%
Val. Loss: 0.700 | Val. Acc: 78.67%

Epoch: 05 | Epoch Time: 1m 17s
Train Loss: 0.083 | Train Acc: 97.59%
Val. Loss: 0.851 | Val. Acc: 79.68%



Experiment 18

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 48
num_warmup_steps=0
lr=2e-5

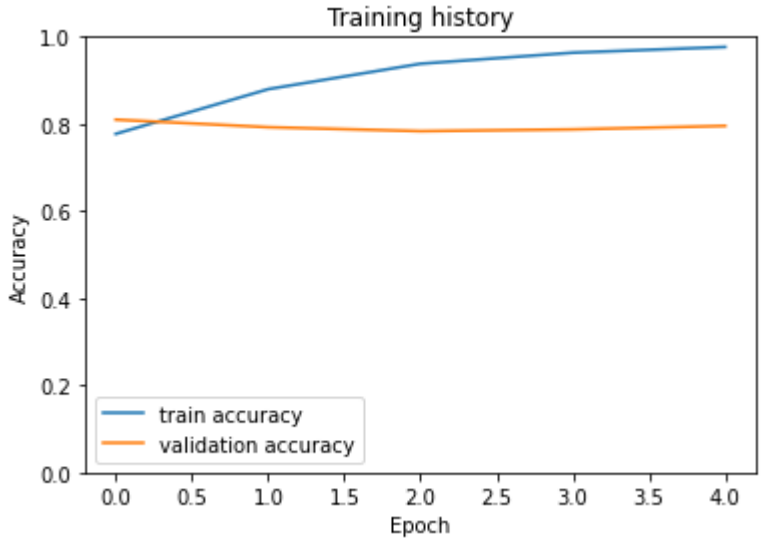
Epoch: 01 | Epoch Time: 1m 11s
Train Loss: 0.443 | Train Acc: 77.71%
Val. Loss: 0.372 | Val. Acc: **80.99%**

Epoch: 02 | Epoch Time: 1m 11s
Train Loss: 0.272 | Train Acc: 87.98%
Val. Loss: 0.417 | Val. Acc: 79.25%

Epoch: 03 | Epoch Time: 1m 11s
Train Loss: 0.163 | Train Acc: 93.82%
Val. Loss: 0.558 | Val. Acc: 78.35%

Epoch: 04 | Epoch Time: 1m 11s
Train Loss: 0.107 | Train Acc: 96.35%
Val. Loss: 0.684 | Val. Acc: 78.75%

Epoch: 05 | Epoch Time: 1m 11s
Train Loss: 0.076 | Train Acc: 97.66%
Val. Loss: 0.854 | Val. Acc: 79.55%



Experiment 19

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 56
num_warmup_steps=0
lr=2e-5

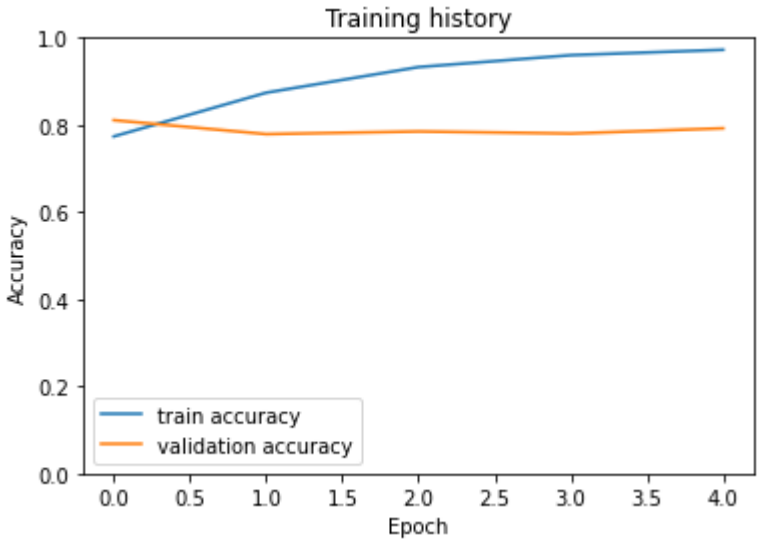
Epoch: 01 | Epoch Time: 1m 8s
Train Loss: 0.450 | Train Acc: 77.36%
Val. Loss: 0.379 | Val. Acc: **81.09%**

Epoch: 02 | Epoch Time: 1m 8s
Train Loss: 0.279 | Train Acc: 87.39%
Val. Loss: 0.416 | Val. Acc: 77.89%

Epoch: 03 | Epoch Time: 1m 8s
Train Loss: 0.168 | Train Acc: 93.28%
Val. Loss: 0.489 | Val. Acc: 78.51%

Epoch: 04 | Epoch Time: 1m 8s
Train Loss: 0.113 | Train Acc: 96.01%
Val. Loss: 0.672 | Val. Acc: 78.03%

Epoch: 05 | Epoch Time: 1m 8s
Train Loss: 0.087 | Train Acc: 97.25%
Val. Loss: 0.789 | Val. Acc: 79.23%



Experiment 20

Length of Queries = 32
Number of Epochs = 5
Pretrained Model = bert-base-uncased
Batch size = 64
num_warmup_steps=0
lr=2e-5

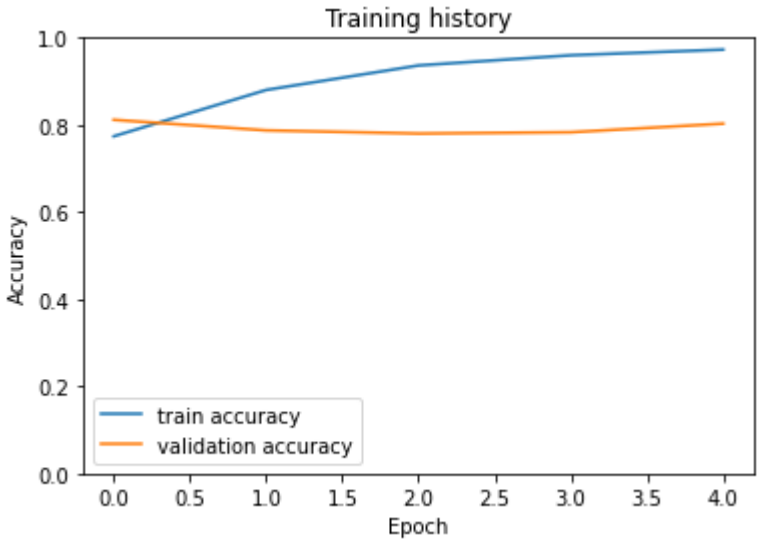
Epoch: 01 | Epoch Time: 1m 8s
Train Loss: 0.451 | Train Acc: 77.39%
Val. Loss: 0.370 | Val. Acc: **81.17%**

Epoch: 02 | Epoch Time: 1m 8s
Train Loss: 0.265 | Train Acc: 88.02%
Val. Loss: 0.418 | Val. Acc: 78.75%

Epoch: 03 | Epoch Time: 1m 8s
Train Loss: 0.166 | Train Acc: 93.65%
Val. Loss: 0.505 | Val. Acc: 78.03%

Epoch: 04 | Epoch Time: 1m 8s
Train Loss: 0.115 | Train Acc: 95.99%
Val. Loss: 0.642 | Val. Acc: 78.32%

Epoch: 05 | Epoch Time: 1m 8s
Train Loss: 0.084 | Train Acc: 97.29%
Val. Loss: 0.730 | Val. Acc: 80.29%



Experiment 21

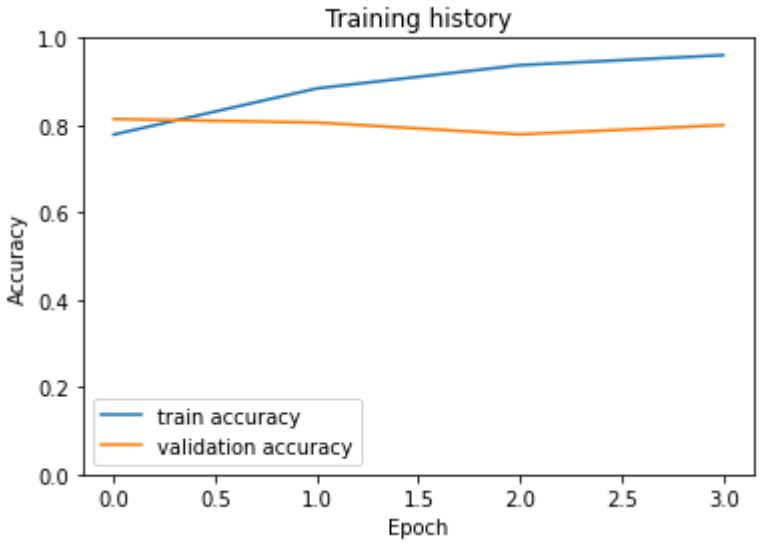
Length of Queries = 32
Number of Epochs = 4
Pretrained Model = bert-base-uncased
Batch size = 32
num_warmup_steps=0
lr=2e-5

Epoch: 01 | Epoch Time: 1m 13s
Train Loss: 0.437 | Train Acc: 77.86%
Val. Loss: 0.371 | Val. Acc: **81.41%**

Epoch: 02 | Epoch Time: 1m 13s
Train Loss: 0.266 | Train Acc: 88.43%
Val. Loss: 0.411 | Val. Acc: 80.61%

Epoch: 03 | Epoch Time: 1m 13s
Train Loss: 0.168 | Train Acc: 93.76%
Val. Loss: 0.528 | Val. Acc: 77.89%

Epoch: 04 | Epoch Time: 1m 13s
Train Loss: 0.117 | Train Acc: 96.05%
Val. Loss: 0.723 | Val. Acc: 80.03%



Experiment 22

Length of Queries = 32
Number of Epochs = 6
Pretrained Model = bert-base-uncased
Batch size = 32
num_warmup_steps=0
lr=2e-5

Epoch: 01 | Epoch Time: 1m 13s
Train Loss: 0.438 | Train Acc: 77.90%
Val. Loss: 0.372 | Val. Acc: **81.31%**

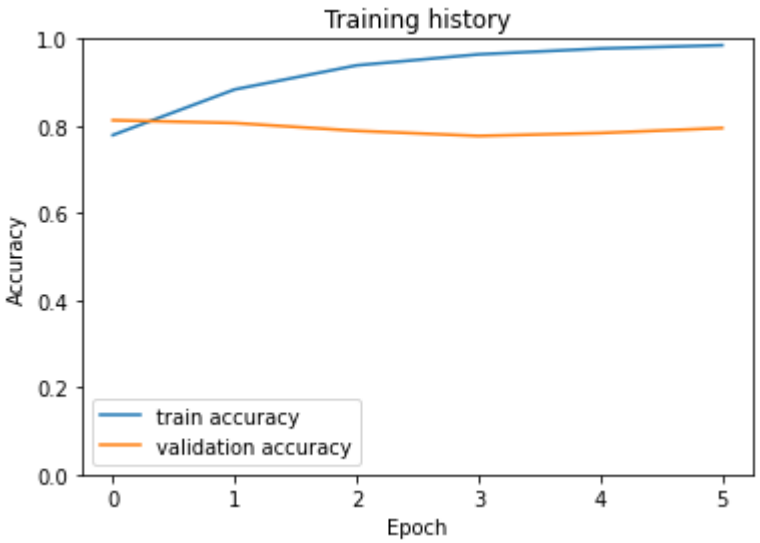
Epoch: 02 | Epoch Time: 1m 13s
Train Loss: 0.268 | Train Acc: 88.38%
Val. Loss: 0.434 | Val. Acc: 80.67%

Epoch: 03 | Epoch Time: 1m 13s
Train Loss: 0.168 | Train Acc: 93.90%
Val. Loss: 0.518 | Val. Acc: 78.88%

Epoch: 04 | Epoch Time: 1m 13s
Train Loss: 0.112 | Train Acc: 96.43%
Val. Loss: 0.724 | Val. Acc: 77.71%

Epoch: 05 | Epoch Time: 1m 13s
Train Loss: 0.075 | Train Acc: 97.76%
Val. Loss: 0.931 | Val. Acc: 78.37%

Epoch: 06 | Epoch Time: 1m 13s
Train Loss: 0.057 | Train Acc: 98.53%
Val. Loss: 1.016 | Val. Acc: 79.52%



Experiment 23

Length of Queries = 32
Number of Epochs = 8
Pretrained Model = bert-base-uncased
Batch size = 32
num_warmup_steps=0
lr=2e-5

Epoch: 01 | Epoch Time: 1m 13s
Train Loss: 0.438 | Train Acc: 77.79%
Val. Loss: 0.373 | Val. Acc: **80.85%**

Epoch: 02 | Epoch Time: 1m 13s
Train Loss: 0.271 | Train Acc: 88.22%
Val. Loss: 0.425 | Val. Acc: 80.69%

Epoch: 03 | Epoch Time: 1m 13s
Train Loss: 0.170 | Train Acc: 93.75%
Val. Loss: 0.494 | Val. Acc: 79.57%

Epoch: 04 | Epoch Time: 1m 13s
Train Loss: 0.113 | Train Acc: 96.29%
Val. Loss: 0.632 | Val. Acc: 78.61%

Epoch: 05 | Epoch Time: 1m 13s
Train Loss: 0.075 | Train Acc: 97.83%
Val. Loss: 0.911 | Val. Acc: 77.97%

Epoch: 06 | Epoch Time: 1m 13s
Train Loss: 0.057 | Train Acc: 98.44%
Val. Loss: 0.996 | Val. Acc: 78.61%

Epoch: 07 | Epoch Time: 1m 13s
Train Loss: 0.038 | Train Acc: 98.97%
Val. Loss: 1.063 | Val. Acc: 79.97%

Epoch: 08 | Epoch Time: 1m 13s
Train Loss: 0.030 | Train Acc: 99.25%
Val. Loss: 1.167 | Val. Acc: 79.84%

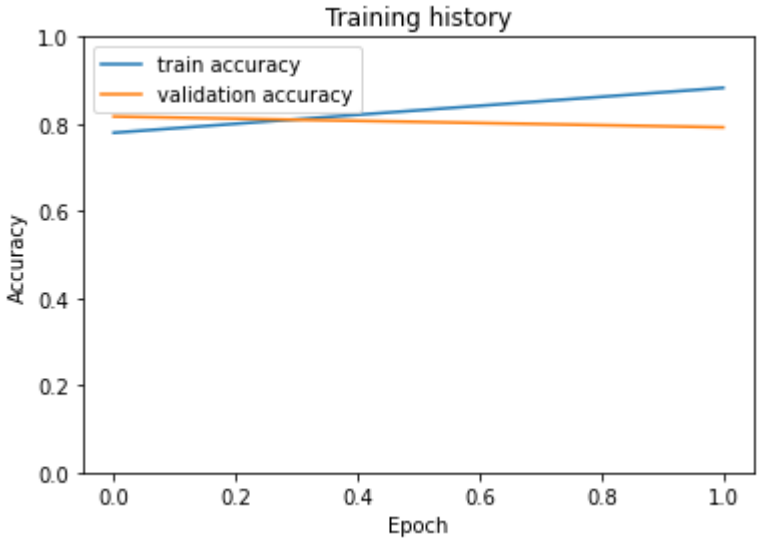
Experiment 24

Length of Queries = 32
Number of Epochs = 2
Pretrained Model = bert-base-uncased
Batch size = 32
num_warmup_steps=0
lr=2e-5

Epoch: 01 | Epoch Time: 1m 13s

Train Loss: 0.435 | Train Acc: 77.98%
Val. Loss: 0.370 | Val. Acc: **81.71%**

Epoch: 02 | Epoch Time: 1m 13s
Train Loss: 0.265 | Train Acc: 88.29%
Val. Loss: 0.424 | Val. Acc: 79.23%



Experiment 25 with Retraining

```
***** Running training *****
  Num examples = 10929
  Num Epochs = 7
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 2394
```

[2394/2394 05:02, Epoch 7/7]

Step	Training Loss
500	2.608500
1000	2.282000
1500	2.056000
2000	1.891500

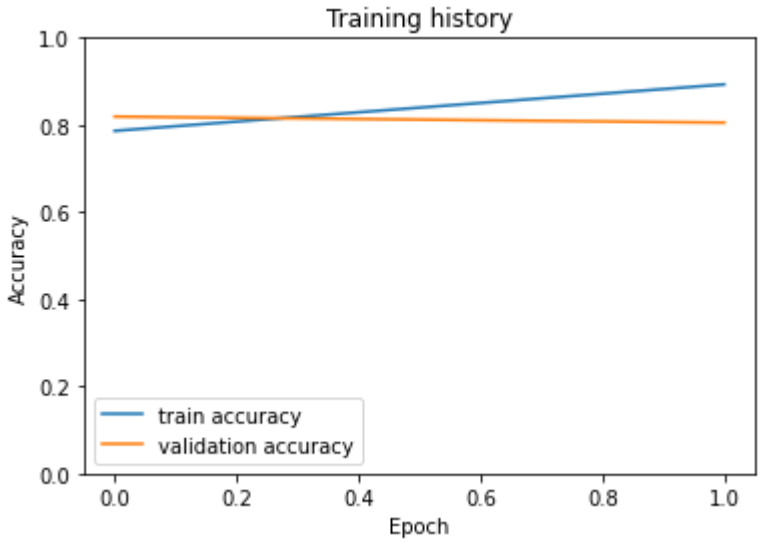
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.429 | Train Acc: 78.65%
Val. Loss: 0.367 | Val. Acc: **81.92%**

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.246 | Train Acc: 89.30%
Val. Loss: 0.418 | Val. Acc: 80.53%



Experiment 26 with Retraining

```
***** Running training *****
  Num examples = 10929
  Num Epochs = 2
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 684
```

[684/684 01:26, Epoch 2/2]

Step	Training Loss
500	2.591000

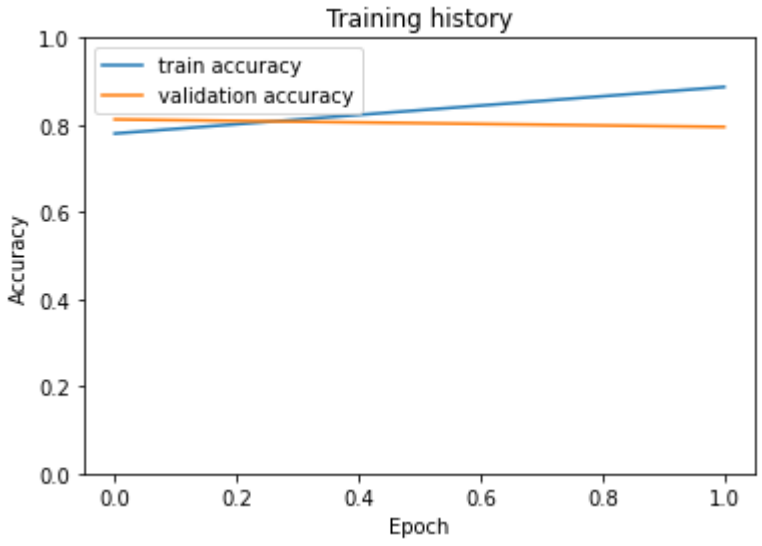
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.431 | Train Acc: 78.03%
Val. Loss: 0.369 | Val. Acc: **81.28%**

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.258 | Train Acc: 88.72%
Val. Loss: 0.420 | Val. Acc: 79.55%



Experiment 27 with Retraining

```
***** Running training *****
  Num examples = 10929
  Num Epochs = 5
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 1710
```

[1710/1710 03:38, Epoch 5/5]

Step	Training Loss
500	2.605000
1000	2.274100
1500	2.056100

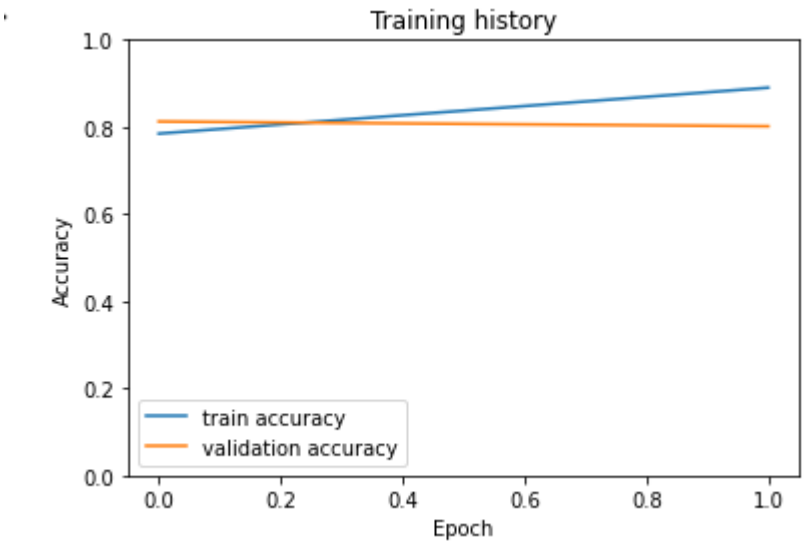
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.425 | Train Acc: 78.47%
Val. Loss: 0.363 | Val. Acc: **81.25%**

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.249 | Train Acc: 89.01%
Val. Loss: 0.415 | Val. Acc: 80.16%



Experiment 28 with Retraining

```
***** Running training *****
  Num examples = 10929
  Num Epochs = 7
  Instantaneous batch size per device = 16
  Total train batch size (w. parallel, distributed & accumulation) = 16
  Gradient Accumulation steps = 1
  Total optimization steps = 4788
```

[4788/4788 07:33, Epoch 7/7]

Step	Training Loss
500	2.763100
1000	2.530900
1500	2.346400
2000	2.191100
2500	2.023600
3000	1.943200
3500	1.948400
4000	1.798300
4500	1.799600

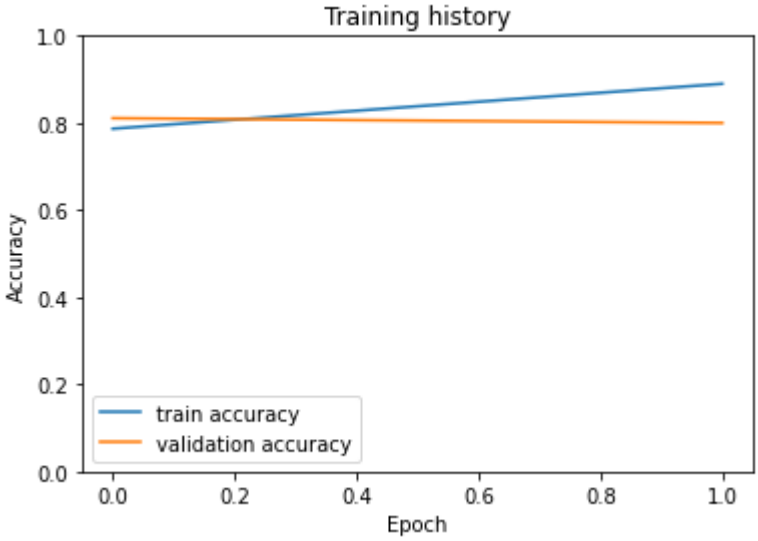
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.422 | Train Acc: 78.67%
Val. Loss: 0.371 | Val. Acc: 81.09%

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.250 | Train Acc: 88.99%
Val. Loss: 0.425 | Val. Acc: 80.00%


Training history
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
train accuracy
validation accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Epoch

Experiment 29 with Retraining

```
***** Running training *****
  Num examples = 10929
  Num Epochs = 7
  Instantaneous batch size per device = 48
  Total train batch size (w. parallel, distributed & accumulation) = 48
  Gradient Accumulation steps = 1
  Total optimization steps = 1596
```

[1596/1596 04:20, Epoch 7/7]

Step	Training Loss
500	2.507600
1000	2.140500
1500	1.971700

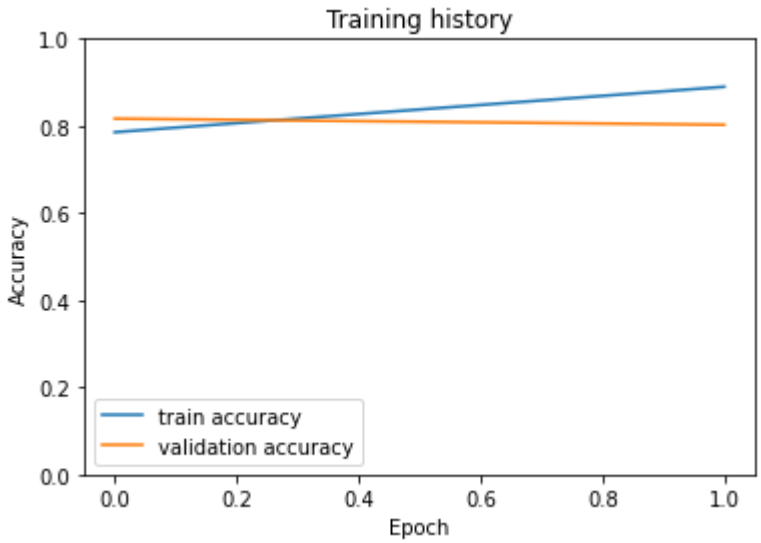
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.424 | Train Acc: 78.57%
Val. Loss: 0.365 | Val. Acc: 81.68%

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.251 | Train Acc: 88.98%
Val. Loss: 0.406 | Val. Acc: 80.27%



Experiment 30 with Retraining

```
***** Running training *****
  Num examples = 17500
  Num Epochs = 5
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 2735
```

[2735/2735 05:51, Epoch 5/5]

Step	Training Loss
500	3.029100
1000	2.648700
1500	2.495000
2000	2.378900
2500	2.320300

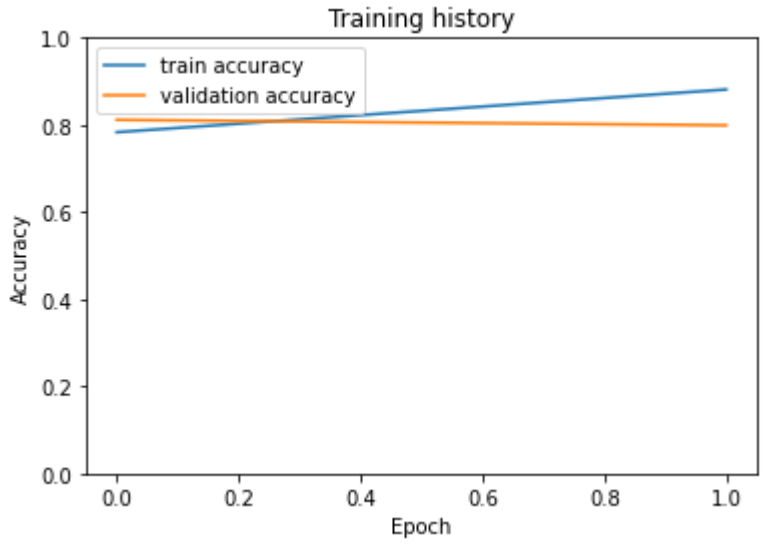
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.431 | Train Acc: 78.33%
Val. Loss: 0.371 | Val. Acc: 81.17%

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.266 | Train Acc: 88.14%
Val. Loss: 0.418 | Val. Acc: 79.92%



Experiment 31 with Retraining

```
***** Running training *****
  Num examples = 17500
  Num Epochs = 7
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 3829
```

[3829/3829 08:11, Epoch 7/7]

Step	Training Loss
500	3.032000
1000	2.656000
1500	2.505300
2000	2.385800
2500	2.310300
3000	2.236600
3500	2.144100

Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

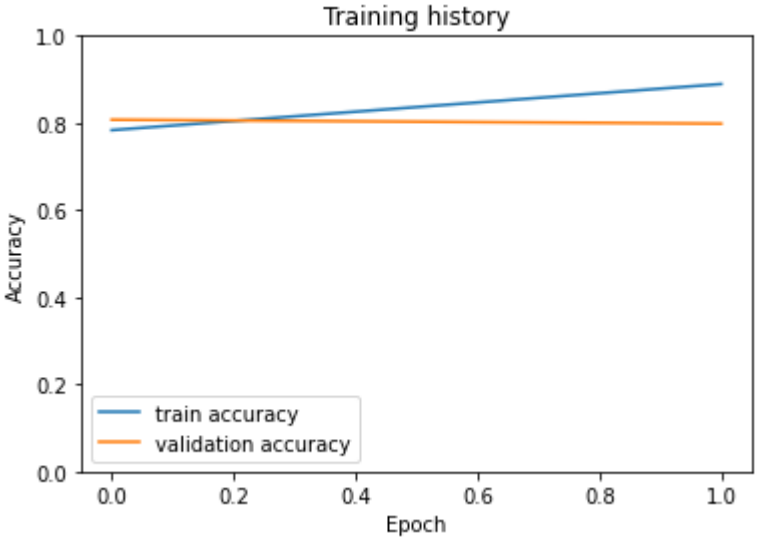
Epoch: 01 | Epoch Time: 1m 16s

Train Loss: 0.430 | Train Acc: 78.35%
Val. Loss: 0.373 | Val. Acc: 80.75%

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.251 | Train Acc: 88.93%

Val. Loss: 0.425 | Val. Acc: 79.84%



Experiment 32 with Retraining

```
***** Running training *****
  Num examples = 17500
  Num Epochs = 2
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 1094
```

[1094/1094 02:19, Epoch 2/2]

Step	Training Loss
500	3.015100
1000	2.625600

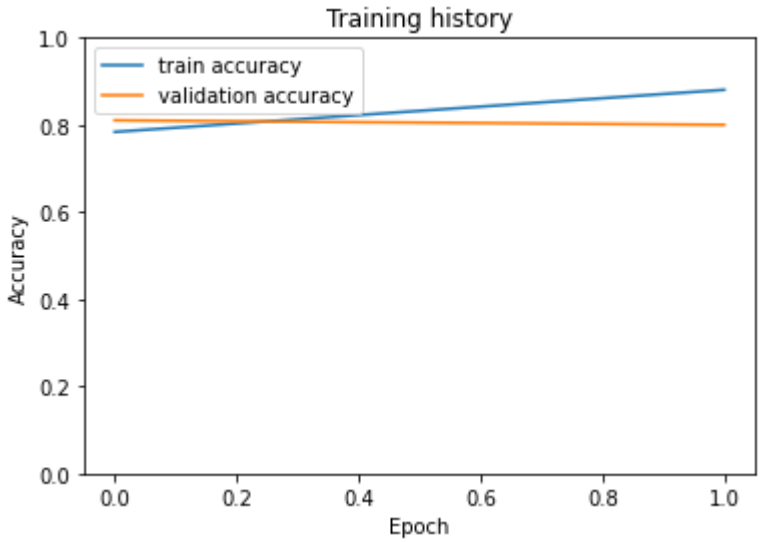
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.432 | Train Acc: 78.40%
Val. Loss: 0.364 | Val. Acc: 81.07%

Epoch: 02 | Epoch Time: 1m 17s

Train Loss: 0.270 | Train Acc: 88.06%
Val. Loss: 0.414 | Val. Acc: 80.03%



Experiment 33 with Retraining

```
***** Running training *****
  Num examples = 6821
  Num Epochs = 5
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 1070
```

[1070/1070 02:16, Epoch 5/5]

Step	Training Loss
500	2.295300
1000	1.921900

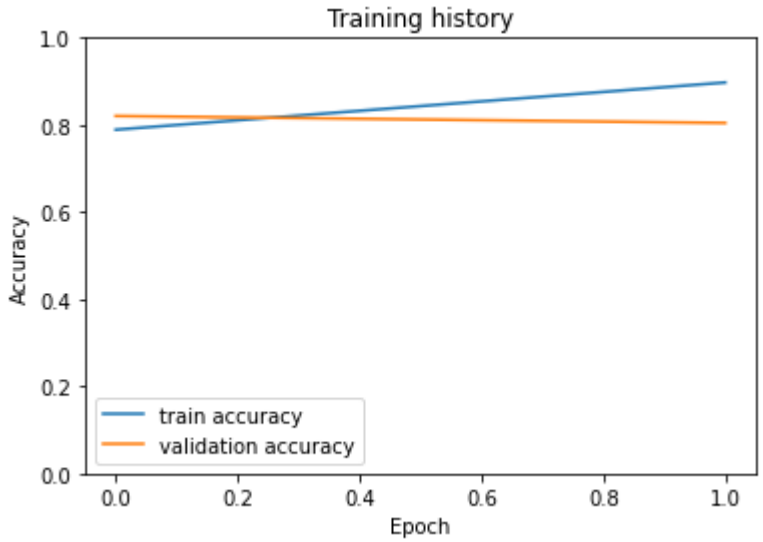
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s

Train Loss: 0.423 | Train Acc: 78.90%
Val. Loss: 0.366 | Val. Acc: 82.05%

Epoch: 02 | Epoch Time: 1m 18s

Train Loss: 0.241 | Train Acc: 89.74%
Val. Loss: 0.417 | Val. Acc: 80.45%



Experiment 34 with Retraining

```
***** Running training *****
  Num examples = 6821
  Num Epochs = 7
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 1498
```

[1498/1498 03:08, Epoch 7/7]

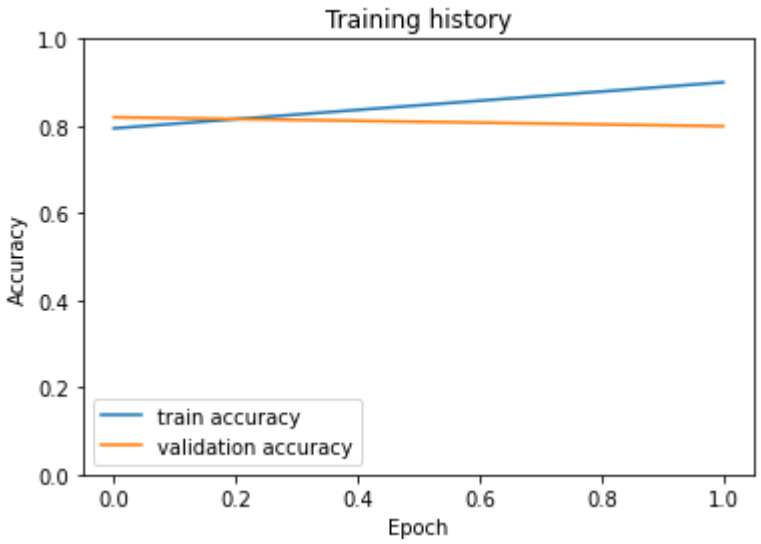
Step	Training Loss
500	2.300000
1000	1.916300

Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 17s
Train Loss: 0.415 | Train Acc: 79.45%
Val. Loss: 0.367 | Val. Acc: 82.00%

Epoch: 02 | Epoch Time: 1m 18s

Train Loss: 0.234 | Train Acc: 90.01%
Val. Loss: 0.429 | Val. Acc: 79.95%



Experiment 35 with Retraining

```
***** Running training *****
  Num examples = 6821
  Num Epochs = 2
  Instantaneous batch size per device = 32
  Total train batch size (w. parallel, distributed & accumulation) = 32
  Gradient Accumulation steps = 1
  Total optimization steps = 428
```

[428/428 00:51, Epoch 2/2]

Step	Training Loss

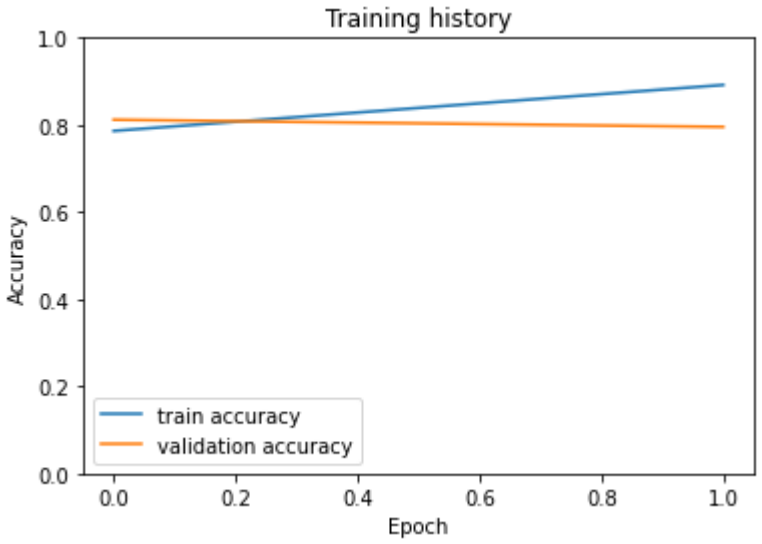
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 18s

Train Loss: 0.425 | Train Acc: 78.63%
Val. Loss: 0.368 | Val. Acc: 81.20%

Epoch: 02 | Epoch Time: 1m 18s

Train Loss: 0.252 | Train Acc: 89.18%
Val. Loss: 0.421 | Val. Acc: 79.55%



Experiment 36 with Retraining

Retraining is same as that in Experiment 33- only change is the parameter tuning during the FineTuning

Length of Queries = 32, Number of Epochs = 5, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 18s

Train Loss: 0.421 | Train Acc: 78.61%
Val. Loss: 0.378 | Val. Acc: 80.59%

Epoch: 02 | Epoch Time: 1m 18s

Train Loss: 0.233 | Train Acc: 90.04%
Val. Loss: 0.437 | Val. Acc: 80.45%

Epoch: 03 | Epoch Time: 1m 17s

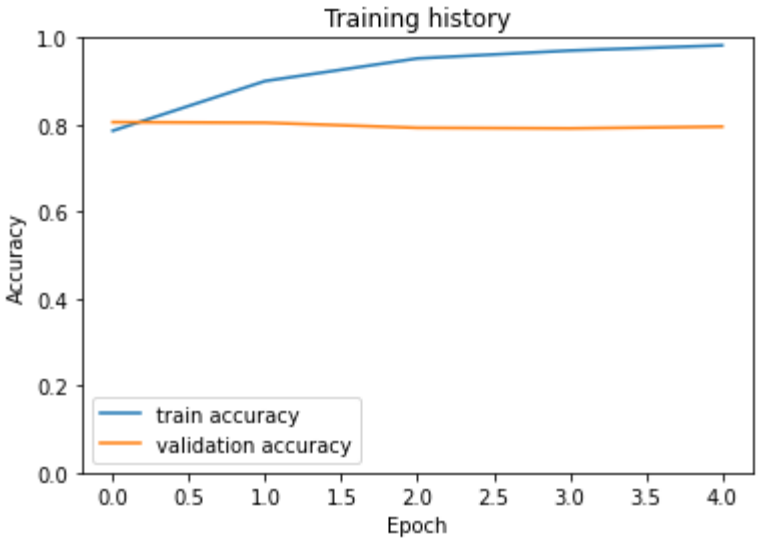
Train Loss: 0.132 | Train Acc: 95.24%
Val. Loss: 0.604 | Val. Acc: 79.28%

Epoch: 04 | Epoch Time: 1m 18s

Train Loss: 0.091 | Train Acc: 97.04%
Val. Loss: 0.816 | Val. Acc: 79.15%

Epoch: 05 | Epoch Time: 1m 17s

Train Loss: 0.060 | Train Acc: 98.25%
Val. Loss: 0.965 | Val. Acc: 79.55%



Experiment 37 with Retraining

Retraining is same as that in Experiment 33- only change is the parameter tuning during the FineTuning- Length of Queries = 32, Number of Epochs = 7, Pretrained Model = bert-base-uncased, Batch size = 32, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 18s

Train Loss: 0.422 | Train Acc: 78.55%
Val. Loss: 0.374 | Val. Acc: 81.33%

Epoch: 02 | Epoch Time: 1m 18s

Train Loss: 0.243 | Train Acc: 89.69%
Val. Loss: 0.458 | Val. Acc: 80.80%

Epoch: 03 | Epoch Time: 1m 18s

Train Loss: 0.140 | Train Acc: 95.05%
Val. Loss: 0.591 | Val. Acc: 80.67%

Epoch: 04 | Epoch Time: 1m 18s

Train Loss: 0.091 | Train Acc: 97.21%
Val. Loss: 0.795 | Val. Acc: 78.85%

Epoch: 05 | Epoch Time: 1m 18s

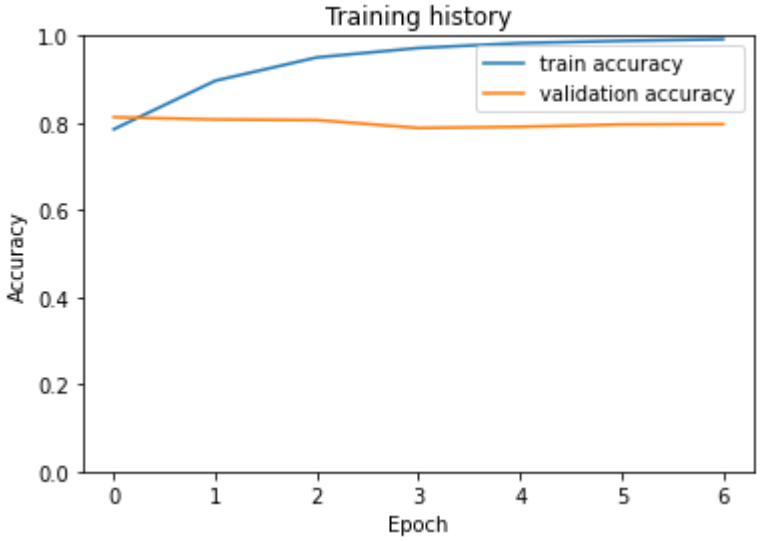
Train Loss: 0.058 | Train Acc: 98.33%
Val. Loss: 0.918 | Val. Acc: 79.12%

Epoch: 06 | Epoch Time: 1m 18s

Train Loss: 0.043 | Train Acc: 98.83%
Val. Loss: 1.086 | Val. Acc: 79.65%

Epoch: 07 | Epoch Time: 1m 18s

Train Loss: 0.030 | Train Acc: 99.22%
Val. Loss: 1.153 | Val. Acc: 79.73%


Training history
train accuracy
validation accuracy
Accuracy
Epoch
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6

Experiment 38 with Retraining

Retraining is same as that in Experiment 33- only change is the parameter tuning during the FineTuning

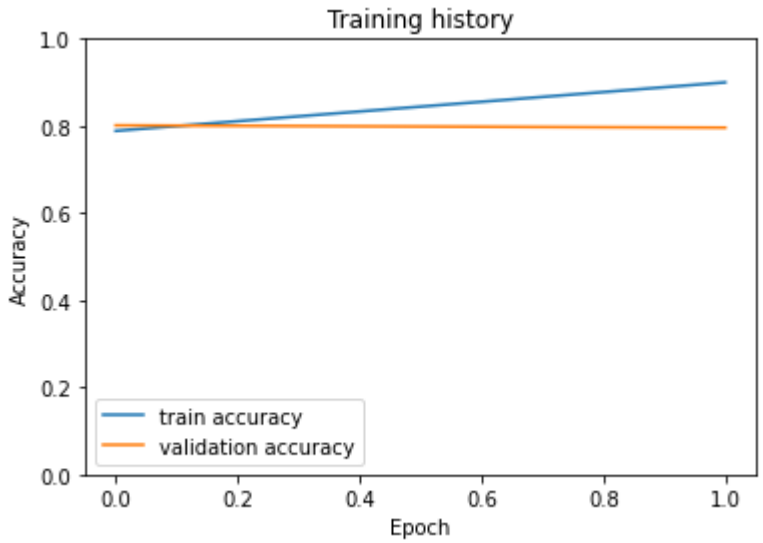
Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 16, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 35s

Train Loss: 0.420 | Train Acc: 78.86%
Val. Loss: 0.377 | Val. Acc: 80.13%

Epoch: 02 | Epoch Time: 1m 36s

Train Loss: 0.242 | Train Acc: 90.00%
Val. Loss: 0.458 | Val. Acc: 79.60%



Experiment 39 with Retraining

Retraining is same as that in Experiment 33- only change is the parameter tuning during the FineTuning

Length of Queries = 32, Number of Epochs = 2, Pretrained Model = bert-base-uncased, Batch size = 48, num_warmup_steps=0, lr=2e-5

Epoch: 01 | Epoch Time: 1m 11s

Train Loss: 0.425 | Train Acc: 78.50%
Val. Loss: 0.374 | Val. Acc: 80.96%

Epoch: 02 | Epoch Time: 1m 11s

Train Loss: 0.247 | Train Acc: 89.01%
Val. Loss: 0.421 | Val. Acc: 79.33%