



+ Code + Text Copy to Drive

Connect ▾

Editing



+ Code

+ Text



```
# run this cell, then restart the runtime before continuing
!pip install git+https://github.com/joeddav/transformers.git@data-collator-type-fix
!pip install nlp
```

```
[ ] from transformers import BertForSequenceClassification, BertTokenizerFast, Trainer, TrainingArguments
    from nlp import load_dataset
    import torch
    import numpy as np
    from sklearn.metrics import accuracy_score, precision_recall_fscore_support
```

```
[ ] model = BertForSequenceClassification.from_pretrained('bert-base-uncased')
    tokenizer = BertTokenizerFast.from_pretrained('bert-base-uncased')
```



Downloading: 100%  433/433 [00:15<00:00, 28.6B/s]

Downloading: 100%  440M/440M [00:14<00:00, 29.5MB/s]

Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertForSequenceClassification: ['cls']
This IS expected if you are initializing BertForSequenceClassification from the checkpoint of a model trained on another task (e.g., language modeling) with a pretrained classifier head (e.g., for text classification).
This IS NOT expected if you are initializing BertForSequenceClassification from the checkpoint of a model that you expect to be initialized with weights for the task (e.g., initialized with weights for text classification).

Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are new to this model. You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

Downloading: 100%  232k/232k [00:01<00:00, 226kB/s]

```
[ ] def tokenize(batch):
```