

2.1 Intro2.2 Processing the data:-

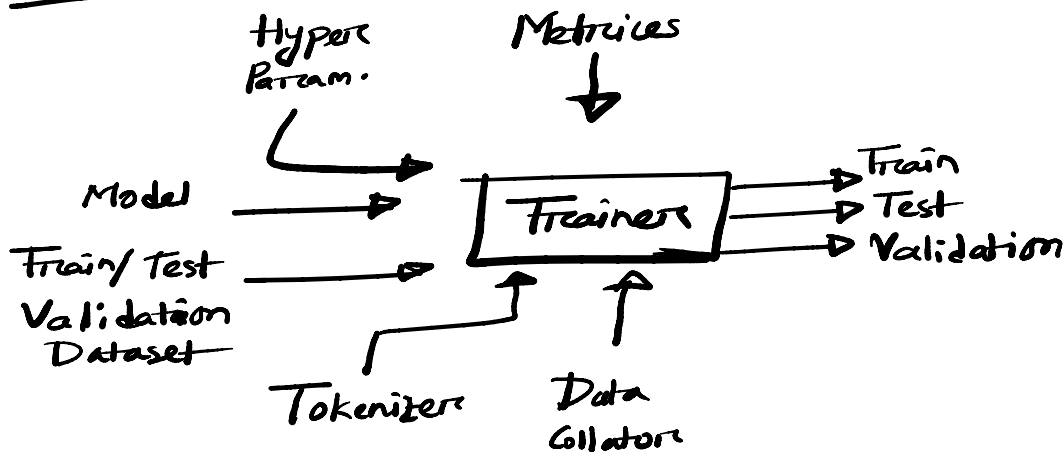
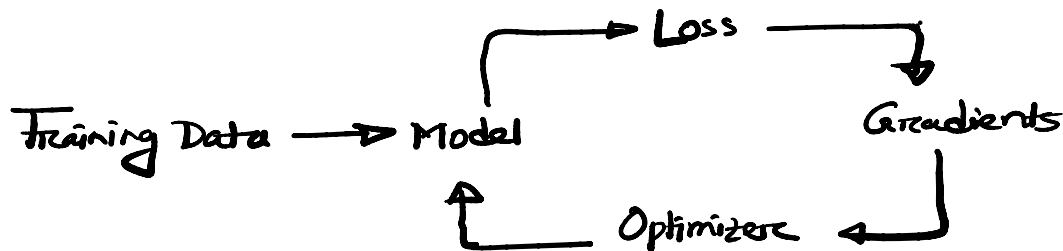
`batch["labels"] = torch.tensor([1, 1])` ?

`Optimizer = AdamW(model.parameters())` : adam optimizer

`loss = model(**batch).loss`

MRPC (Microsoft Research Paraphrase Corpus)

- `remove_columns()`
- `rename_column()`
- `with_format("torch")`

3.3 Trainers3.1 A full training:-

`trainer.predict('validation dataset')`

trainer.predict('validation dataset')

return a named tuple with three fields

1. predictions \rightarrow logits

2. label_ids

3. metrics

3.1 loss

3.2 time

3.2.1 how long to predict
in total/average

AdamW

SGD

modifies weight decay in Adam

Decoupling weight decay from grad. update.

L2 regularization

Where