

Multiple-choice training (e.g. SWAG)

This folder contains the `run_swag.py` script, showing an examples of *multiple-choice answering* with the 😊 Transformers library. For straightforward use-cases you may be able to use these scripts without modification, although we have also included comments in the code to indicate areas that you may need to adapt to your own projects.

Multi-GPU and TPU usage

By default, the script uses a `MirroredStrategy` and will use multiple GPUs effectively if they are available. TPUs can also be used by passing the name of the TPU resource with the `--tpu` argument.

Memory usage and data loading

One thing to note is that all data is loaded into memory in this script. Most multiple-choice datasets are small enough that this is not an issue, but if you have a very large dataset you will need to modify the script to handle data streaming. This is particularly challenging for TPUs, given the stricter requirements and the sheer volume of data required to keep them fed. A full explanation of all the possible pitfalls is a bit beyond this example script and README, but for more information you can see the 'Input Datasets' section of [this document](#).

Example command

```
python run_swag.py \
  --model_name_or_path distilbert-base-cased \
  --output_dir output \
  --do_eval \
  --do_train
```