## Environment info

- `transformers` version:
- Platform:
- Python version:
- PyTorch version (GPU?):
- Tensorflow version (GPU?):
- Using GPU in script?:
- Using distributed or parallel set-up in script?:

**Who can help**

## Information

Model I am using (Bert, XLNet . . . ):

The problem arises when using: * [ ] the official example scripts: (give details below) * [ ] my own modified scripts: (give details below)

The tasks I am working on is: * [ ] an official GLUE/SQUaD task: (give the name) * [ ] my own task or dataset: (give details below)

## To reproduce

Steps to reproduce the behavior:

1.
2.
3.

## Expected behavior