

Modified from DeepLearning.AI Hugging Face course

<https://learn.deeplearning.ai/courses/open-source-models-hugging-face/lesson/3/natural-language-processing-nlp>

```
!pip install transformers
!pip install torch
!pip install sentence-transformers
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.38.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.12.2)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.20.3)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (24.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.5.15)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.32.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.19.1)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.4.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.66.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.5.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.11.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.7.4)
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages (2.1.0+cu121)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch) (3.12.2)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from torch) (4.11.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch) (3.3)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from torch) (3.1.3)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch) (2023.5.0)
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch) (2.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2) (2.1.5)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy) (1.3.0)
Collecting sentence-transformers
  Downloading sentence_transformers-2.5.1-py3-none-any.whl (156 kB)
```

156.5/156.5 kB 3.4 MB/s eta 0:00:00

```
Requirement already satisfied: transformers<5.0.0,>=4.32.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (4.38.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (4.66.1)
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (2.1.0+cu121)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.26.4)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.5.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.12.0)
Requirement already satisfied: huggingface-hub>=0.15.1 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (0.20.3)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (10.3.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (3.12.2)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (2024.5.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (2.32.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (4.11.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (24.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch) (3.3)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages (from torch) (3.1.3)
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages (from torch) (2.1.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.5.15)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.19.1)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.4.3)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.5.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2) (2.1.5)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.7.4)
```

```

# suppress warnings
from transformers.utils import logging
logging.set_verbosity_error()

from sentence_transformers import SentenceTransformer

model = SentenceTransformer("all-MiniLM-L6-v2")

/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.com/settings/tokens).
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or dataset
  warnings.warn(
modules.json: 100% 349/349 [00:00<00:00, 12.9kB/s]
config_sentence_transformers.json: 100% 116/116 [00:00<00:00, 3.66kB/s]
README.md: 100% 10.7k/10.7k [00:00<00:00, 492kB/s]
sentence_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 2.31kB/s]
config.json: 100% 612/612 [00:00<00:00, 24.4kB/s]
pytorch_model.bin: 100% 90.9M/90.9M [00:00<00:00, 115MB/s]
/usr/local/lib/python3.10/dist-packages/torch/_utils.py:831: UserWarning: TypedStorage is deprecated
  return self.fget.__get__(instance, owner)()
tokenizer_config.json: 100% 350/350 [00:00<00:00, 10.9kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 2.55MB/s]
tokenizer.json: 100% 466k/466k [00:00<00:00, 4.53MB/s]
special_tokens_map.json: 100% 112/112 [00:00<00:00, 6.39kB/s]
1_Pooling/config.json: 100% 190/190 [00:00<00:00, 9.33kB/s]

sentences1 = ['The cat sits outside',
              'A man is playing guitar',
              'The movies are awesome']

embeddings1 = model.encode(sentences1, convert_to_tensor=True)

embeddings1
tensor([[ 0.1392,  0.0030,  0.0470, ...,  0.0641, -0.0163,  0.0636],
        [ 0.0227, -0.0014, -0.0056, ..., -0.0225,  0.0846, -0.0283],
        [-0.1043, -0.0628,  0.0093, ...,  0.0020,  0.0653, -0.0150]])

sentences2 = ['The dog plays in the garden',
              'A woman watches TV',
              'The new movie is so great']

embeddings2 = model.encode(sentences2,
                           convert_to_tensor=True)

print(embeddings2)

```

```
tensor([[ 0.0163, -0.0700,  0.0384, ...,  0.0447,  0.0254, -0.0023],
        [ 0.0054, -0.0920,  0.0140, ...,  0.0167, -0.0086, -0.0424],
        [-0.0842, -0.0592, -0.0010, ..., -0.0157,  0.0764,  0.0389]])
```

```
from sentence_transformers import util
```

```
cosine_scores = util.cos_sim(embeddings1, embeddings2)
```

```
# the first line compares similarity to the 3 sentences in sentences_2, etc
print(cosine_scores)
```

```
tensor([[ 0.2838,  0.1310, -0.0029],
        [ 0.2277, -0.0327, -0.0136],
        [-0.0124, -0.0465,  0.6571]])
```

```
for i in range(len(sentences1)):
    print("{} \t\t {} \t\t Score: {:.4f}".format(sentences1[i],
                                                sentences2[i],
                                                cosine_scores[i][i]))
```

The cat sits outside	The dog plays in the garden	Score: 0.2838
A man is playing guitar	A woman watches TV	Score: -0.0327
The movies are awesome	The new movie is so great	Score: 0.6571