BERT

Solving tasks with BERT

Learning goals

- defined the key learning goals here
- second learning goal

HUGGINGFACE

- General Link: `https://huggingface.co/`
- `transformers`: Python library for state-of-the-art NLP
  ▸ Wolf et al., 2020 ▸ huggingface transformers documentation
- Further I: Model Hub for sharing model weights
  ▸ huggingface models
- Further II: Hosting APIs / Apps for interacting with models
  ▸ huggingface spaces
- Further III: ▸ `datasets` module that is perfectly integrated with their `transformers` library

DOCUMENT CLASSIFICATION

- Assume the ▸ IMDB data set:

```
from datasets import load_dataset
dataset = load_dataset("imdb", split = "train")
```

- Split the training data into train and validation

```
train_data = dataset.train_test_split(test_size = 0.2)
```

```
DatasetDict({
    train: Dataset({
        features: ['text', 'label'],
        num_rows: 20000
    })
    test: Dataset({
        features: ['text', 'label'],
        num_rows: 5000
    })
})
```

DOCUMENT CLASSIFICATION

- Load model and tokenizer:

```
1 tokenizer = AutoTokenizer.from_pretrained("bert-base-
      uncased", use_fast=True)
2 model = AutoModelForSequenceClassification.from_pretrained
      ("bert-base-uncased", num_labels=2)
```

A SLIDE WITH A FIGURE