

## Question Answering

- ✓ **Video:** Week 3 Overview  
6 min
- ✓ **Reading:** Week 3 Overview  
10 min
- ✓ **Video:** Transfer Learning in NLP  
7 min
- 📖 **Reading:** Transfer Learning in NLP  
10 min
- 📖 **Video:** ELMo, GPT, BERT, T5  
7 min
- 📖 **Reading:** ELMo, GPT, BERT, T5  
10 min
- 📖 **Video:** Bidirectional Encoder Representations from Transformers (BERT)  
4 min
- 📖 **Reading:** Bidirectional Encoder Representations from Transformers (BERT)  
10 min
- 📖 **Video:** BERT Objective  
2 min
- 📖 **Reading:** BERT Objective  
10 min
- 📖 **Video:** Fine tuning BERT  
2 min
- 📖 **Reading:** Fine tuning BERT  
10 min
- 📖 **Video:** Transformer: T5  
3 min
- 📖 **Reading:** Transformer T5  
10 min
- 📖 **Video:** Multi-Task Training Strategy  
5 min
- 📖 **Reading:** Multi-Task Training Strategy  
10 min
- 📖 **Video:** GLUE Benchmark  
2 min
- 📖 **Reading:** GLUE Benchmark  
10 min
- 📖 **Video:** Question Answering  
2 min
- 📖 **Reading:** Question Answering  
10 min
- 📖 **Lab:** SentencePiece and BPE  
2h
- 📖 **Reading:** Content Resource  
10 min

## Assignment

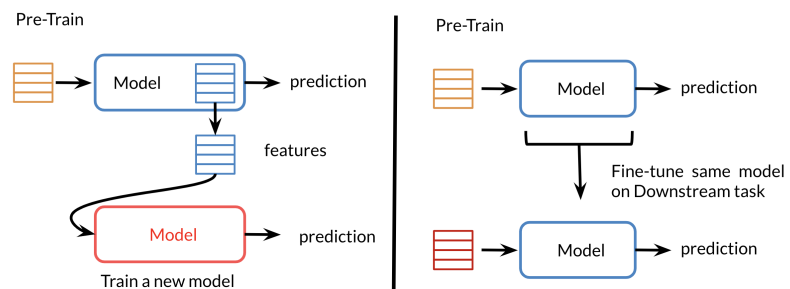
# Transfer Learning in NLP

There are three main advantages to transfer learning:

- Reduce training time
- Improve predictions
- Allows you to use smaller datasets

Two methods that you can use for transfer learning are the following:

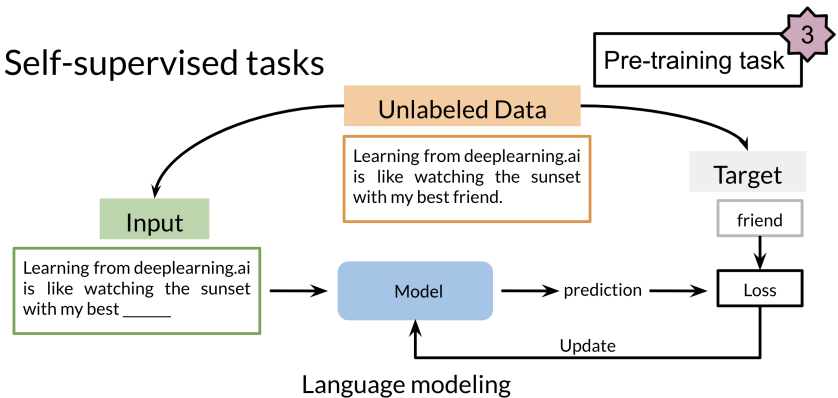
## Feature-based vs. Fine-Tuning



In feature based, you can train word embeddings by running a different model and then using those features (i.e. word vectors) on a different task.

When fine tuning, you can use the exact same model and just run it on a different task. Sometimes when fine tuning, you can keep the model weights fixed and just add a new layer that you will train. Other times you can slowly unfreeze the layers one at a time. You can also use unlabelled data when pre-training, by masking words and trying to predict which word was masked.

## Self-supervised tasks



For example, in the drawing above we try to predict the word "friend". This allows your model to get a grasp of the overall structure of the data and to help the model learn some relationships within the words of a sentence.

Mark as completed