8.6

Large Language Models

Part 1: The Two Main Ingredients

Sebastian Raschka and the Lightning AI Team

Original transformer ("Attention is All You Need") was developed for language translation

Today, transformers (large language models) are also used for ...

- Classification (e.g., BERT)

- Various text summarization and generation tasks (e.g., GPT)

- Conversational chatbots (e.g., ChatGPT)

- Protein structure prediction from sequence data (e.g., AlphaFold 2)

The 2 main ingredients that make transformers so successful

Self-attention (Unit 8.5)

Self-supervised pretraining (Unit 7.7)
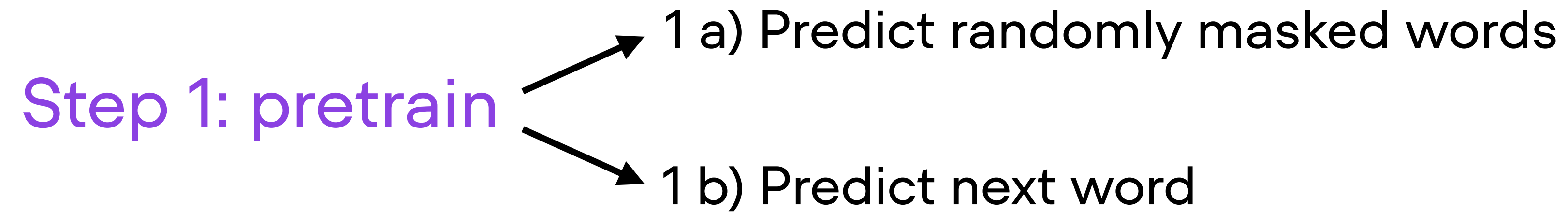
Self-supervised pretraining

Step 1: pretrain on large unlabeled dataset
(learn a general language model)

Self-supervised pretraining

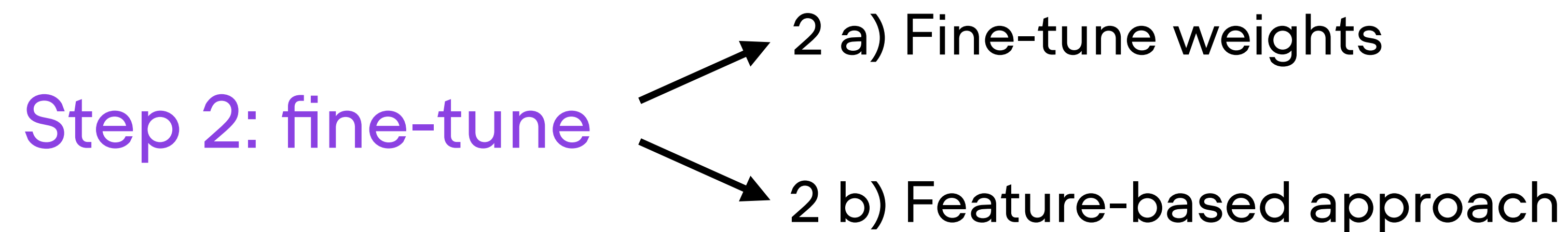Step 1: pretrain on large unlabeled dataset
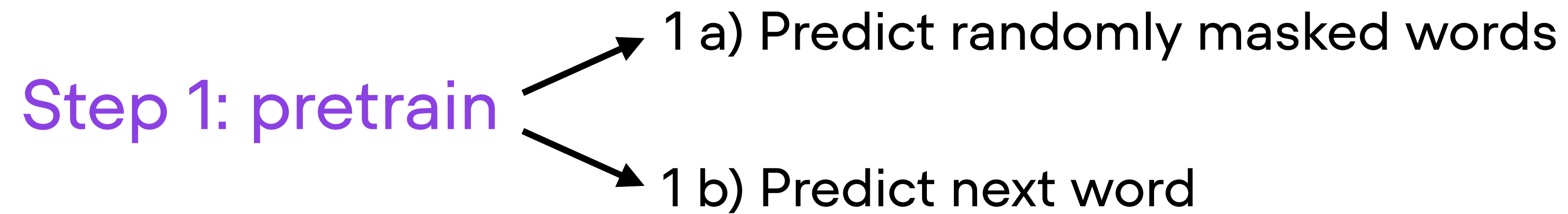(learn a general language model)

Step 2: fine-tune on smaller, labeled target dataset

Self-supervised pretraining

Step 1: pretrain

1 a) Predict randomly masked words

1 b) Predict next word

Step 2: fine-tune

Self-supervised pretraining

Step 1: pretrain

1 a) Predict randomly masked words

1 b) Predict next word

Step 2: fine-tune

2 a) Fine-tune weights

2 b) Feature-based approach

Next: Pretraining in GPT