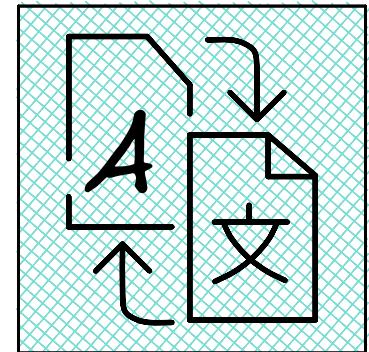
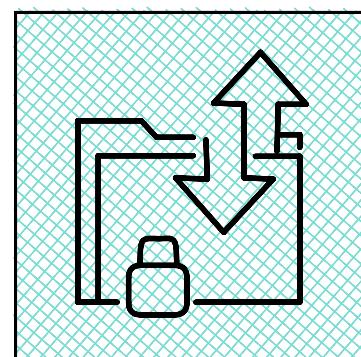


Transfer Learning

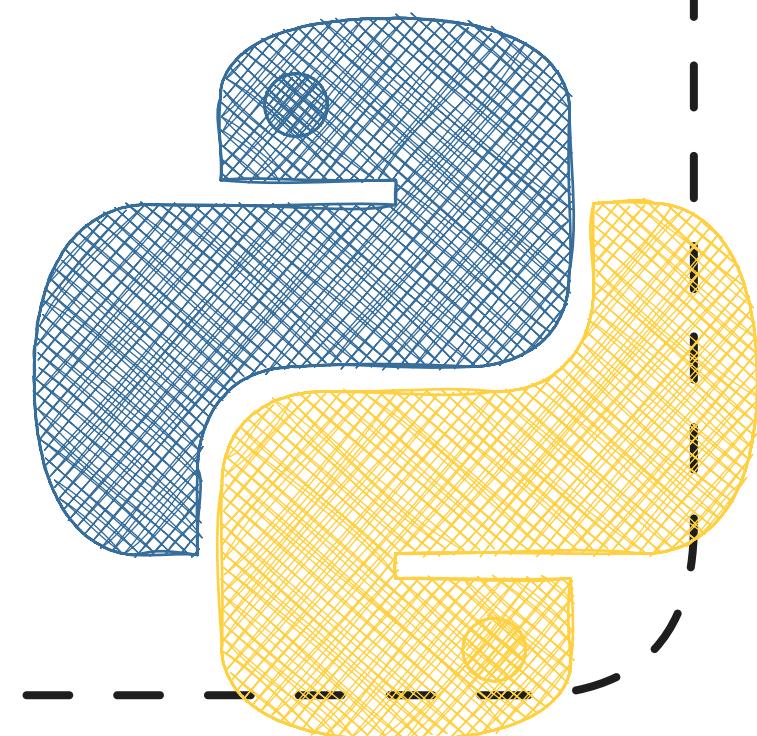
visão intuitiva



Técnica que aproveita o conhecimento previamente adquirido

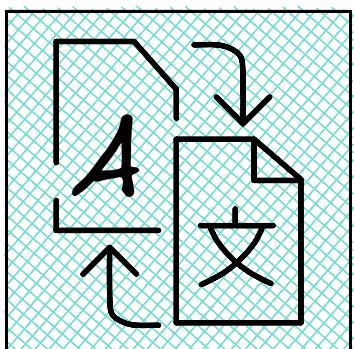


Transfer



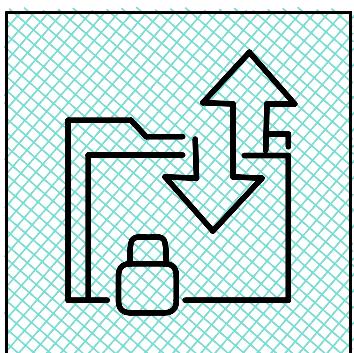
Reutilização direta

Feature Extraction



Um modelo pré-treinado é usado como extrator fixo de embeddings

Principal característica

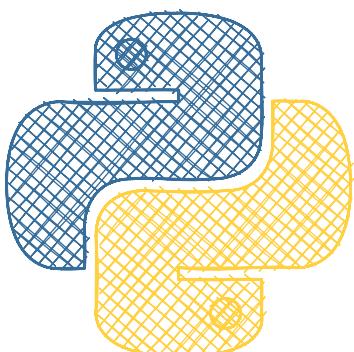


Camadas de convolução permanecem fixas

Apenas o classificador final é treinado

Exemplo:

usar a ResNet50 pré-treinada no ImageNet e treinar apenas uma camada densa que classifica diferentes tipos de folhas



Novo contexto

Fine-Tuning

Parte ou todo o modelo pré-treinado é reajustado

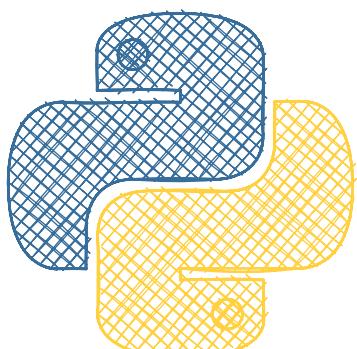
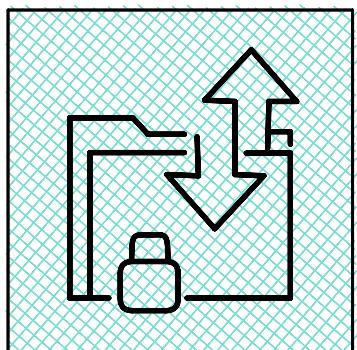
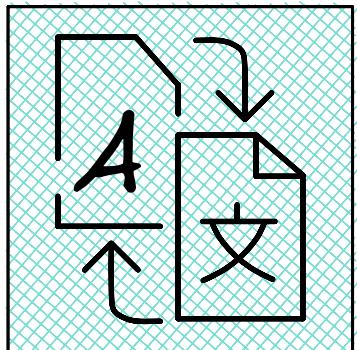
Principal característica

Permite que o modelo se adapte ao novo domínio

É mantido o conhecimento geral aprendido

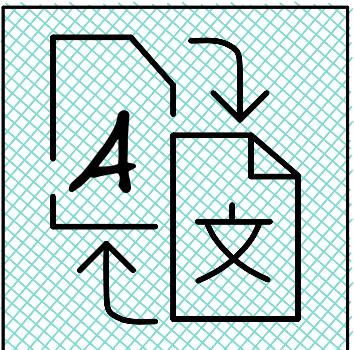
Exemplo:

ajustar o BERT para análise de sentimentos em texto em português



Compressão de conhecimento

Knowledge Distillation

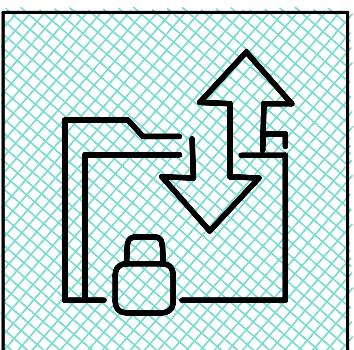


Transfere o conhecimento de um modelo grande para um modelo pequeno

Principal característica



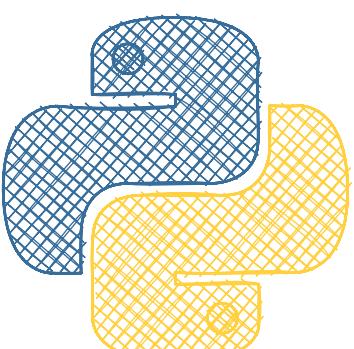
O aluno aprende a imitar as previsões do professor



Mantendo desempenho semelhante com menos parâmetros

Exemplo:

distilar o GPT-4 em um modelo para rodar em dispositivos móveis



Principais formas de Transfer Learning

Categoria / Nome	Descrição	Exemplo prático
Fine-tuning	Re-treinar o modelo inteiro ou parcial em uma nova tarefa	Fine-tune GPT-2 em texto jurídico
Feature extraction / Frozen model	Usar embeddings pré-treinados e treinar só um classificador ou novo head	Usar BERT embeddings para classificação
Distillation	Treinar um modelo pequeno ("student") a partir de um modelo grande ("teacher")	Distilar GPT-4 para um GPT-2 menor
Prompt tuning / Soft prompts	Aprender um vetor de prompt (embeddings) — não muda o modelo	P-tuning, Prompt Tuning
Adapter tuning (Adapters)	Inserir camadas pequenas (adapters) no modelo — só treinar essas camadas	Adapter-BERT, PEFT Adapters
LoRA (Low-Rank Adaptation)	Treinar matrizes de baixa-rank em camadas do modelo (parâmetros muito reduzidos)	LoRA fine-tuning em LLaMA
Prefix tuning	Aprender um "prefixo" de embeddings que guia o modelo — sem alterar o core do modelo	Prefix-tuning GPT-2
BitFit	Só ajustar os bias (bias-only fine-tuning)	BitFit para tuning de BERT
Continual pre-training (Domain-adaptive pre-training)	Continuar o pre-training em novo domínio antes do fine-tuning	Pretrain BERT em textos médicos antes de fine-tuning para NER
Multi-task learning	Treinar em múltiplas tarefas simultaneamente (transferência entre tarefas)	T5 treinado para tradução + summarization
Cross-lingual transfer	Usar um modelo treinado em um idioma em outro idioma	Usar mBERT para português sem re-treinar
Zero-shot transfer	Aplicar o modelo em uma nova tarefa apenas com prompting	Usar GPT-4 para gerar SQL sem fine-tuning
Few-shot learning	Fornecer alguns exemplos no prompt ("in-context learning")	GPT-4 few-shot learning

Source

- [Hugging Face Course Transfer Learning](#)