Homework 1

1. 使用 GPTQ 量化 OPT-6.7B 模型。课程代码:

<https://github.com/DjangoPeng/LLM-quickstart/blob/main/transformers/AutoGPTQ_transformers.ipynb>

1. 使用 AWQ 量化 Facebook OPT-6.7B 模型。Facebook OPT 模型地址：

<https://huggingface.co/facebook?search_models=opt>

课程代码：

<https://github.com/DjangoPeng/LLM-quickstart/blob/main/transformers/AWQ_transformers.ipynb>

Homework 2

1. 在“LoRA 低秩适配 OpenAI Whisper-Large-V2 语音识别任务”中，为中文语料的训练过程增加过程评估，观察 Train Loss 和 Validation Loss 变化。

课程代码:

<https://github.com/DjangoPeng/LLM-quickstart/blob/main/peft/peft_lora_whisper-large-v2.ipynb>

1. 在“LoRA 低秩适配 OpenAI Whisper-Large-V2 语音识别任务”中，当 LoRA 模型训练完成后，使用测试集进行完整的模型评估。

课程代码:

<https://github.com/DjangoPeng/LLM-quickstart/blob/main/peft/peft_lora_whisper-large-v2.ipynb>