

LLM HW1 Report

Hosna Oyarhoseini - 402212503

Table1: Accuracy for each method:

Note book	Full	Soft prompt			Adapte r-1	Adapter-2		LoRA		
			N_SOFT_P ROMPT_T OKENS=1	N_SOFT_PR OMPT_TOKE NS=10		BOTTLEN ECK_SIZ E =8	BOTTLENE CK_SIZE =1		RANK=1	RANK=8
Accuracy%	90.0	87.1	84.8	86.6	90.6	90.4		89.5	88.8	88.9

Table2: number of parameters for each method:

Fine-tuning method	Full fine-tuning	Soft prompt	Adapter	LoRA
Number of parameters	60,506,624	60,511,744	60,611,168	60,801,536
Trained parameters	60,506,624	5120	104,544	294,912

Reasoning behind the number of parameters trained in each method:

- **Full fine-tuning:**

We train **all** the parameters of model again, which is **60,506,624**

- **Soft prompt:**

T5 has 512 hidden units we add 10 tokens to each and train only them in finetuning so we would have **$512 \times 10 = 5120$** parameters which trained during soft prompt fine tuning

- **Adapter:**

T5-small has 12 blocks (6 blocks for encoder and 6 blocks for decoder) each block needs an adapter and we use bottle neck with size 8, so each of our adapter has:

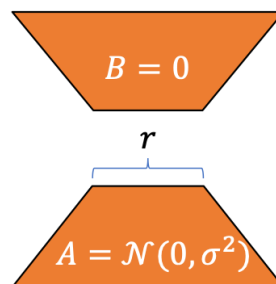
$$\begin{aligned}
 (512 + 1) * 8 &\quad \longrightarrow (\text{input} + \text{bias}) * \text{hidden layer} \\
 + \\
 (8 + 1) * 512 &\quad \longrightarrow (\text{hidden layer} + \text{bias}) * \text{output} \\
 = 8,712 \text{ parameters}
 \end{aligned}$$

So for the whole model we would have:

$$12 * 8712 = 104,544 \text{ parameters}$$

- LoRA:

T-5 small has 18 layers (6 block for encoder each have 1 layer, and 6 block for decoder each with 2 layer) each has one key and value, we use a network like a picture below for each of them:



Where input and output are 512 and $r = \text{RANK} = 8$ so each of these networks has: $512 * 8$ parameters.

So for each key or value in each layer, we have $(8 * 512) + (8 * 512)$ parameters

Thus for each layer we have:

$$2 * 8 * 512 * 2 \text{ parameters}$$

And for the whole network we added:

$$18 * 2 * 2 * 8 * 512 \text{ parameters}$$

To finetune using LoRA, which means **294,912** parameters trained.