

به نام خدا

گزارش تمرین اول درس مدل‌های زبانی بزرگ

حمیدرضا امیرزاده

۴۰۱۲۰۶۹۹۹

الف) دقت روش های مختلف تنظیم دقیق:

Approach (#Notebook)	Max val. accuracy after 10 epochs (%)
Full Fine-Tuning (#1)	90.0
Soft Prompt Tuning_10 (#2)	86.8
Soft Prompt Tuning_10_OpenDelta (#2)	87.1
Soft Prompt Tuning_1_OpenDelta (#2)	84.9
Adapter_8 (#3)	90.6
Adapter_8_AdapterHub (#4)	90.3
Adapter_1_AdapterHub (#4)	89.1
LoRA_8 (#5)	88.9
LoRA_8_PEFT (#5)	89.0
LoRA_1_PEFT (#5)	88.8

توجه:

- عدد نشان داده شده بعد از روش های Soft Prompt Tuning برابر تعداد توکن های اضافی یا N_SOFT_PROMPT_TOKENS می باشد.
- عدد نشان داده شده بعد از روش های Adapter برابر اندازه گلوگاه آداپتور یا BOTTLENECK_SIZE می باشد.
- عدد نشان داده شده بعد از روش های LoRA برابر رنک ماتریس های A و B یا rank می باشد.

الف) تعداد پارامتر روش های مختلف تنظیم دقیق:

Approach (#Notebook)	Total number of parameters
Full Fine-Tuning (#1)	60506624
Soft Prompt Tuning_10 (#2)	60511744
Adapter_8 (#3)	60611168
LoRA_8 (#5)	60801536

- در روش **Soft Prompt Tuning_10** که ۱۰ توکن اضافی وجود دارد و اندازه امبدینگ هر توکن برابر ۵۱۲ می باشد، می توان تعداد پارامترهای اضافه مدل را به صورت زیر محاسبه کرد:

$$\text{Additinal number of params} = \text{number of soft tokens} * \text{embedding size of each token} \\ = 10 * 512 = 5120$$

مشاهده می شود که در جدول بالا نیز تعداد پارامترهای اضافه شده مدل در این روش برابر ۵۱۲۰ می باشد:

$$60511744 - 60506624 = 5120$$

- در روش **Adapter_8** که اندازه گلوگاه آداپتور برابر ۸ است دارد و اندازه امبدینگ هر توکن برابر ۵۱۲ می باشد و این آداپتور به ۱۲ لایه اضافه می شود (۶ لایه انکودر و ۶ لایه دیکودر در مدل T5-small)، می توان تعداد پارامترهای اضافه مدل را به صورت زیر محاسبه کرد:

$$\text{Additinal number of params} = \text{total number Of layers} * (\text{number of params in a single adapter})$$

$$\text{number of params in a single adapter} = (\text{embedding size of each token} * \text{bottleneck_size} * 2) + (\text{embedding size of each token} + \text{bottleneck_size})$$

توجه شود که ترم اول مربوط به تعداد وزن ها و ترم دوم مربوط به تعداد بایاس ها می باشد.

$$\Rightarrow \text{Additinal number of params} = 12 * ((512 * 8 * 2) + (512 + 8)) = 104544$$

مشاهده می شود که در جدول بالا نیز تعداد پارامترهای اضافه شده مدل در این روش برابر ۱۰۴۵۴۴ می باشد:

$$60611168 - 60506624 = 104544$$

- در روش **LoRA_8** که رنگ ماتریس ها برابر ۸ است و اندازه امبدینگ هر توکن برابر ۵۱۲ می باشد و این ماژول به تمامی ماژول های توجه مربوط به کوئری و ارزش اضافه می شود (۶ ماژول self-attention در انکودر، ۶ ماژول self-attention و ۶ ماژول cross-attention در دیکودر در مدل T5-small)، می توان تعداد پارامترهای اضافه مدل را به صورت زیر محاسبه کرد:

$$\text{Additinal number of params} = \text{number of encoder layers} * 2 * (\text{embedding size of each token} * \text{rank} * 2) + \text{number of decoder layers} * 2 * 2 * (\text{embedding size of each token} * \text{rank} * 2)$$

$$= 6 * 2 * (512 * 8 * 2) + 6 * 2 * 2 * (512 * 8) = 294912$$

مشاهده می شود که در جدول بالا نیز تعداد پارامترهای اضافه شده مدل در این روش برابر ۲۹۴۹۱۲ می باشد:

$$60801536 - 60506624 = 294912$$