

مدل‌های زبانی بزرگ

پاییز ۱۴۰۲

استاد: دکتر سلیمانی، دکتر رهبان، دکتر عسگری

گردآورندگان: حامد جمشیدیان، زینب سادات تقوی، سید محمدرضا مدرس

بررسی و بازبینی: محمد علی صدرایی جواهری



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

مهلت ارسال: ۱۹ آبان

تنظیم دقیق با بهینگی پارامتر

تمرین اول

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است. برای انجام تمرین زمان کافی اختصاص داده شده است. انجام آن را به هیچ وجه به روزهای پایانی موکول نکنید.
- سوالات خود را فقط از طریق **کوئرا** درس و در نوشته‌ی مربوط به اطلاع‌رسانی این تمرین بپرسید.
- حتما در نام‌گذاری فایل‌های آپلودی خود از قالب $\{Name\}_{STD_Number}$ تبعیت کنید.
- در طول ترم ۵ روز تاخیر مجاز برای ارسال تکالیف دارید. پیشنهاد می‌شود تاخیرهای خود را برای مواقع ضروری نگه دارید.
- پاسخ‌های ارسالی باید منحصرا حاصل تلاش فردی شما باشد. در صورت استفاده از منابع خارجی یا همفکری، حتما این موارد را ذکر کنید. همچنین توصیه می‌شود **آداب نامه‌ی انجام تمرین‌های درسی** را مطالعه کنید. برای اطلاع از قوانین خاص این درس به فایل قوانین درس بر روی کوئرا مراجعه کنید.

توضیحات (۱۰۰ نمره)

روش‌های تنظیم دقیق با بهینگی پارامتر (Parameter-Efficient Fine-Tuning) که به طور مخفف با نام PEFT شناخته می‌شوند، مجموعه روش‌هایی هستند که به جای تغییر تمام پارامترهای مدل، تعداد اندکی از آن‌ها را موقع یادگیری وظیفه جدید تغییر می‌دهند. این پارامترهای در حال تغییر می‌توانند زیر مجموعه‌ای از پارامترهای اصلی مدل باشند یا پارامترهای جدیدی باشند که به مدل اضافه می‌شوند. این روش‌ها در ۵ سال گذشته مورد توجه محققان قرار گرفته‌اند و سه مزیت کلی می‌توان برای آن‌ها ذکر کرد:

۱. بخشی از میزان رم مصرفی در هنگام آموزش مدل مخصوص Optimizer است مقدار رمی که Optimizer مصرف می‌کند وابسته به تعداد پارامترهایی است که قرار است آموزش ببینند. لذا استفاده از روش‌های PEFT مقدار رم مصرفی را کاهش می‌دهد.
 ۲. بعد از اتمام آموزش مدل نیاز به ذخیره تمام پارامترهای مدل نیست. این موضوع در هنگامی پر کاربرد است که قرار باشد یک مدل بنیادین را بر روی وظایف مختلف آموزش دهید. در صورت استفاده از PEFT در ازای هر وظیفه مقدار حافظه کمی باید استفاده شود، چون لازم نیست پارامترهای اصلی مدل نگهداری شود.
 ۳. در زمان تنظیم دقیق کردن مدل از پیش آموزش دیده همیشه خطر overfit برای مدل وجود دارد. اما در PEFT چون تمام پارامترهای مدل تغییر نمی‌کند، دانش قبلی مدل آسیب نمی‌بیند و این خطر برای مدل کمتر است.
- سه کتابخانه اصلی این روش‌ها را پیاده سازی کرده‌اند که در این تمرین از هر سه آنها قرار است استفاده شود.

۱. PEFT

۲. AdapterHub

۳. OpenDelta

شما باید یکی از مدل‌های بنیادین با نام T5-small را بر روی دیتاست imdb با ۴ روش متفاوت fine-tune کنید. برای این کار نوتبوک‌های از پیش آماده شده‌ای در اختیار شما قرار گرفته‌اند. روش‌های مورد نیاز در زیر لیست شده‌اند:

۱. Full Fine-Tuning

۲. Soft Prompt Tuning

۳. Adapter

۴. LoRA

نوتبوک‌های تکمیل شده باید کنار گزارش آپلود شوند. در مجموع ۵ نوتبوک در اختیار شما قرار گرفته است (روش Adapter دارای دو نوتبوک است). برای هر روش PEFT باید مساله را در دو حالت زیر حل کنید:

۱. تکمیل کد پیاده‌سازی از ابتدا

۲. تکمیل کد استفاده از کتابخانه

کتابخانه‌ای که برای هر روش مجاز به استفاده از آن هستید در همان نوتبوک مشخص شده است.

شما باید در نهایت ۱۰ عدد را در گزارش خود بیاورید (یک عدد از نوتبوک اول، سه عدد از نوتبوک دوم، یک عدد از نوتبوک سوم، دو عدد از نوتبوک چهارم، سه عدد از نوتبوک پنجم).
وزن‌های PEFT آموزش دیده در روش‌های غیر کتابخانه‌ای باید آپلود شوند. که شامل سه فایل زیر می‌شود:

۱. lora.pt

۲. adapter.pt

۳. prompts.pt

در بخش آخر گزارش خود با یک جدول تعداد پارامترهای هر ۴ روش در حالت غیر کتابخانه‌ای را گزارش کنید. این تعداد با نام Number of parameters در نوتبوک‌ها چاپ شده‌اند. در گزارش کار خود برای هر روش PEFT محاسبه کنید که بر اساس هایپرپارامترهای آن روش چه تعداد پارامتر به مدل اضافه شده است. اعداد حساب شده را با اعداد جدول مذکور مقایسه کنید. محاسبات شما باید حاصل یکسانی با تعداد پارامتر گزارش داده شده در جدول داشته باشند.

در نهایت یک فایل PDF گزارش و دو فایل زیپ برای این تمرین باید آپلود شود که در محل آپلود در کویرا توضیحات لازم هر کدام آمده است. در نهایت توزیع نمره تمرین به شکل زیر است. ۵ درصد نمره بیشتر برای ارفاق به دانشجویان در نظر گرفته شده است.

۱. نوتبوک اول ۵%

۲. نوتبوک دوم ۲۵%

۳. نوتبوک سوم ۲۵%

۴. نوتبوک چهارم ۵%

۵. نوتبوک پنجم ۳۰%

۶. گزارش ۱۵%