



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



یادگیری عمیق با کاربرد در تصویر و صوت

دکتر رشاد حسینی

تمرین شماره ۴

طرح تمرین:

فاطمه نادی

خرداد ۱۴۰۴

مقدمه.....	۶
سوال ۱.....	۷
سوال ۲ مراحل توسعه مدل های زبانی بزرگ.....	۹
سوال ۳ پیاده سازی.....	۱۱
مراجع.....	۱۴
نکات تحویل.....	۱۵
پیاده سازی.....	۱۵
گزارش.....	۱۶
بارگذاری.....	۱۷
ارتباط با تدریسار.....	۱۷

فهرست اشکال

- شکل ۱: زیرشبکه رمزگشا (راست) و رمزگذار (چپ) در معماری مبدل [۱] ۷
- شکل ۲: ساختار هد تکی توجه [۱] ۹

فهرست جداول

- جدول ۱: ابعاد مورد نیاز برای محاسبه بعد و وزنهای معماری مبذل ۸
- جدول ۲: ابرپارامترهای پیشنهادی برای آموزش مدل ۱۲

جدول اصطلاحات تخصصی

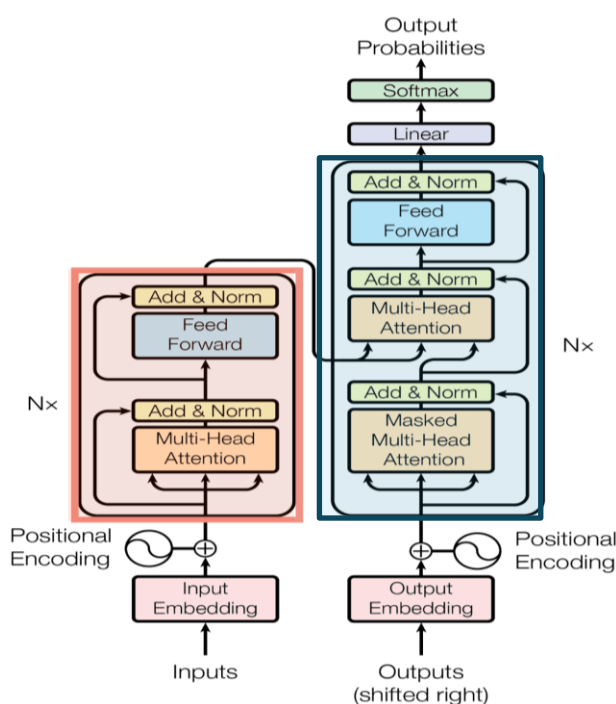
اصطلاح	مخفف	معادل فارسی
Batch		دسته
Transformer model		مدل مبدل
Encoder- Decoder		رمزنگار-رمزگشا
Representation		ستون فقرات پیچشی
Learning rate		نرخ یادگیری
Multi head Attention		هد توجه چندگانه
Feed forward		پیش‌رو
Masked Multi-Head Attention		هد چندگانه ماسک‌شده توجه
Pre-training		پیش آموزش
Post-training		پس آموزش
Alignment		هم‌راستا سازی
Token embedding		بردار تعبیه توکن
Positional embedding		بردار تعبیه موقعیت مکانی

در این تمرین قصد داریم به جهت آشنایی بیشتر شما با مدل‌های متنی در ابتدا شما را با نحوه‌ی عملکرد مدل‌های مبدل آشنا کرده و در نهایت با پیاده‌سازی یک مدل زبانی کوچک به طور عملی با نحوه آموزش این مدل‌ها آشنایی پیدا کنید.

سوال ۱

در این سوال قصد داریم تا با عملکرد مدل‌های مبدل کمی دقیق‌تر آشنا شویم. این مدل از مکانیزم توجه استفاده می‌کند که اولین بار در سال ۲۰۱۷ معرفی شد [۱] و به‌طور معمول در تحقیقات پردازش زبان‌های طبیعی استفاده می‌شود.

معماری مبدل از ساختار رمزگذار-رمزگشا پیروی می‌کنند. وظیفه رمزگذار (در سمت چپ شکل ۱)، نگاشت دنباله ورودی به دنباله‌ای از بازنمایی‌های میانی است، که سپس به یک رمزگشا وارد می‌شود. رمزگشا (در سمت راست شکل ۱)، خروجی رمزگذار را همراه با خروجی رمزگشا در مرحله زمانی قبلی دریافت می‌کند تا یک دنباله خروجی ایجاد کند.



شکل ۱: زیرشبکه رمزگشا (راست) و رمزگذار (چپ) در معماری مبدل [۱]

در مقاله‌ی اصلی تعداد لایه‌های رمزگذار و رمزگشا، ۶ در نظر گرفته شده است که البته می‌توان هر تعدادی را در نظر گرفت. در رمزگذار هر لایه از دو زیر لایه مجزا تشکیل شده است، که لایه‌ی اول هد چندگانه توجه و لایه‌ی دوم یک شبکه پیشرو خطی است. رمزگشا شباهت‌های زیادی با رمزگذار دارد. این شبکه نیز از پشته‌ای از N لایه یکسان تشکیل شده است که هر یک از سه زیر لایه مجزا تشکیل شده است. که شامل هد چندگانه ماسک‌شده توجه، هد چندگانه توجه و شبکه پیشرو خطی است.

در این سوال قصد داریم تعداد پارامترها و ابعاد بردارها پس از عبور از لایه‌های مختلف را بررسی کنیم.
 ابرپارامترهای معماری مربوط به مبدل در جدول ۱ آورده شده است.

جدول ۱: ابعاد مورد نیاز برای محاسبه بعد و وزنهای معماری مبدل

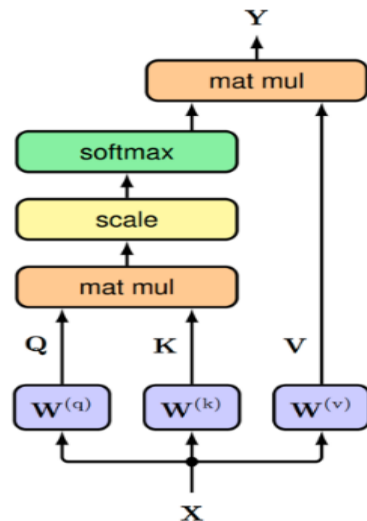
مقدار	ابرپارامتر
۳۲	اندازه بسته
۱۰۰۰	اندازه واژگان
۴۰۹۶	بیشترین طول ورودی
۱۰۲۴	امبدینگ توکن ورودی
۷۶۸	بردار نهان
۶	بلوک انکودر
۸	بلوک دیکدر
۶	هد توجه
۳	لایه‌های پسرو

در لایه پسرو اول، ابعاد ورودی نصف می‌گردد، در لایه دوم بعد تغییری نمی‌کند و در نهایت در لایه سوم بعد دو برابر شده و به همان مقدار بعد اولیه برمی‌گردد.

۱. ابعاد سیگنال را قبل از ورود به هر بلوک و پس از خروج از آن به طور مرحله به مرحله و صرفاً برای بلوک رمزنگار بنویسید. توجه شود برای یکی از هدهای توجه لازم است تا ابعاد به طور دقیق نوشته شود. (راهنمایی: به این منظور با توجه به شکل ۲ ابعاد را در هر بلوک محاسبه کنید) (۱۰ نمره)

۲. پارامترهای این معماری را به تفکیک هر بلوک نوشته و هم برای رمزنگار و هم برای رمزگشا گزارش

دهید. (۱۰ نمره)



شکل ۲: ساختار هد تکی توجه [۱]

سوال ۲ مراحل توسعه مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ هم‌چون GPT و LLaMA طی چندین مرحله آموزش داده می‌شوند تا بتوانند خروجی‌های معنادار، منطقی و هم‌راستا با اهداف انسانی تولید کنند. به‌طور معمول، فرایند آموزش این مدل‌ها در سه مرحله‌ی کلی انجام می‌شود:

۱. پیش آموزش

۲. پس آموزش که خود شامل:

o آموزش با نظارت^۱

o هم‌راستاسازی

شما باید به موارد زیر پاسخ دهید:

۱. مرحله پیش آموزش

الف) این مرحله در چه داده‌هایی و با چه هدفی انجام می‌شود؟ (۳ نمره)

ب) چرا مدل در این مرحله هنوز آماده‌ی استفاده مستقیم در کاربردهای انسانی نیست؟ (۳ نمره)

¹ Supervised Fine-Tuning

۲. مرحله آموزش با نظارت

الف) تفاوت این مرحله با پیش آموزش چیست؟ (۳ نمره)

ب) نوع داده‌هایی که در این مرحله استفاده می‌شوند را شرح دهید و یک مثال بزنید؟ (۳ نمره)

ج) چرا این مرحله نقش مهمی در تعامل انسانی-مدل دارد؟ (۳ نمره)

۳. همراستاسازی

الف) هدف اصلی از این مرحله چیست؟ (۲ نمره)

ب) یکی از روش‌های رایج برای انجام همراستاسازی را نام ببرید و نحوه‌ی کار آن را به طور

خلاصه توضیح دهید. (۳ نمره)

سوال ۳ پیاده‌سازی

در این سوال قصد داریم یک نمونه ابتدایی شده از مدل GPT را به صورت کامل از پایه پیاده‌سازی کنیم.

قسمت یک: مجموعه داده

برای این تمرین از دیتاست زیر استفاده کنید:

`wget https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-08/friends.csv`

دیتاست friends.csv شامل دیالوگ‌های شخصیت‌های اصلی سریال Friends است. ستون‌های مهم این دیتاست شامل موارد زیر است:

- Season : شماره فصل
 - episode : شماره قسمت
 - Speaker : نام شخصیتی که دیالوگ را گفته است
 - text : متن دیالوگ
 - Scene : شماره صحنه داخل هر قسمت
- در این تمرین، فقط از ستون text استفاده می‌کنیم و مدل زبان را روی دیالوگ‌ها آموزش می‌دهیم. هدف این تمرین، پیاده‌سازی یک مدل زبانی ساده با استفاده از مدل‌های مبدل برای تولید دیالوگ جدید بر اساس دیالوگ‌های سریال است.

قسمت دو: تنظیمات اولیه پیشنهادی

شما می‌توانید از پارامترهای زیر برای طراحی و تنظیم مدل خود استفاده نمایید. در صورت استفاده از مقادیر دیگر حتماً آن‌را در گزارش خود بیاورید.

جدول ۲: ابرپارامترهای پیشنهادی برای آموزش مدل

پارامتر	مقدار
<i>batch_size</i>	16
<i>block_size</i>	32
<i>max_iterion</i>	5000
<i>eval_interval</i>	100
<i>learning_rate</i>	$1e - 3$
<i>n_embd</i>	64
<i>n_head</i>	4

قسمت سه: پیاده‌سازی ماژول‌های اصلی مدل (۲۵ نمره)

مدلی طراحی کنید که شامل اجزای زیر باشد:

۱. بردار تعبیه کاراکتر: یک لایه `nn.Embedding` برای نگاشت هر توکن به برداری با ابعاد `n_embd`.

۲. بردار تعبیه موقعیت مکانی: لایه‌ای برای مدل کردن موقعیت هر توکن در دنباله.

۳. بلاک مبدا: پیاده‌سازی یک بلاک ترنسفورمر شامل:

○ Multi Head Self Attention و Self Attention Head

○ Feed Forward Network

○ لایه نرمال‌سازی و اتصال‌های باقی‌مانده^۲

۴. نرمال‌سازی نهایی: لایه نرمال‌سازی پس از عبور از بلاک‌ها.

۵. لایه خروجی خطی: برای تولید لاجیت^۳ بر روی فضای واژگان.

^۲ residual

^۳ logit

قسمت چهار: آموزش مدل (۱۵ نمره)

تابع `train_model` را پیاده‌سازی کنید و مدل را آموزش دهید. سپس در بازه‌های مشخص، مقدار تابع هزینه را گزارش کند. انتخاب بازه‌ها به عهده دانشجو می‌باشد.

قسمت پنج: تولید متن (۲۰ نمره)

تابعی با نام `generate_text` پیاده‌سازی کنید و با دادن یک توکن اولیه، دنباله‌ای از توکن‌ها با استفاده از مدل تولید کند. چند نمونه از خروجی‌های مدل را گزارش کرده و نتایج را تحلیل کنید.

- [1] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

پیاده‌سازی

- به منظور مدیریت و نگهداری منبع کد پیاده‌سازی، ساختاربندی آن باید به صورت درختواره زیر باشد؛ توضیح مربوط به کارکرد هر فایل و پوشه آورده شده است .

homework_code/

```
|— data/          # Data-related files
| |— data_loader.py # Data loading and preprocessing
| |— dataset/      # Raw data (if applicable)
|— models/        # Model-related files
| |— model.py      # Define your neural network architecture
| |— saved_models/ # Save trained models
|— notebooks/     # Jupyter notebooks (optional)
| |— exploratory.ipynb # Initial data exploration
|— scripts/       # Python scripts
| |— main.py       # Main script to run the entire pipeline
| |— train.py      # Training loop
| |— evaluate.py   # Model evaluation
|— config/        # configurations
| |— config.yaml   # Store hyperparameters
| |— logging.yaml  # Logging, Checkpoints (optional)
|— utils/         # Utility functions or helper modules
| |— visualization.py # Plotting functions
| |— metrics.py    # evaluation metrics
|— README.md      # Project documentation (optional)
|— .gitignore     # Git ignore file (optional)
```

- کد شما باید قابلیت اجرا بر روی بخش کوچکی از داده‌ها را داشته باشد تا دستیار آموزشی بتواند در مدت زمان کوتاهی پیاده‌سازی شما را ارزیابی کند و توابع و کلاسها توضیح داشته باشند.

- از آدرس‌دهی مطلق در کدهای خود استفاده نکنید و به جای آن از آدرس‌دهی نسبی استفاده نمایید.
- در صورت مشاهده‌ی موارد همانندی بین دو یا چند فرد در گزارشکار و یا کد به طرفین تقلب **نمره** **صفر** داده خواهد شد. به کپی برداری از کدهای آماده اینترنت، تصویر صفحه^۴ یا کد دیگران نمره‌های تعلق نخواهد گرفت.
- اگر برای بخشی از پیاده‌سازی، از کدهای آماده اینترنتی استفاده می‌کنید که جزء قسمت‌های اصلی تمرین نمی‌باشد، حتما باید لینک آن در گزارش و کد ارجاع داده شود. در غیر اینصورت تقلب محسوب شده و کل نمره تمرین را از دست می‌دهید ولی محدودیتی در استفاده از منابع اینترنتی ندارید.

گزارش

- گزارش شما در فرایند تصحیح از اهمیت ویژه‌ای برخوردار خواهد بود. لذا تمامی نکات و فرض‌هایی که برای محاسبات و پیاده‌سازی خود در نظر می‌گیرید را در گزارش ذکر کنید. دقت داشته باشید که **۵۰ درصد** از نمره تمرین شما مربوط به گزارش است.
- الزامی به ارائه توضیح جزئیات کد در گزارش نمی‌باشد. اما باید نتایج بدست آمده را به صورت کامل گزارش و تحلیل کنید.
- گزارش حتما **به فارسی** نوشته و از برابر فارسی لغات استفاده شود.^۵ در صورت تمایل می‌توانید از قالب لیتک^۶ نیز برای نوشتن گزارش استفاده نمایید.
- در گزارش خود برای تصاویر **زیرنویس** و برای جداول **بالانویس** اضافه نمایید و به آنها ارجاع دهید. این توضیحات باید به طور مستقل منظور از شکل یا جدول را برسانند.
- از تصویر صفحه برای گزارش نتیجه یا روابط ریاضی استفاده نشود و نتیجه باید به صورت عددی یا جدول آورده شده و روابط ریاضی نیز نوشته شود.

^۴ Screen Shot

^۵ برای نمونه به جای Epoch، اپاک یا دوره را به کار برده و در صورت نیاز برابر انگلیسی پانویس شود.

^۶ Latex

- اگر شکل از منبعی گرفته شده باید ارجاع داده شود و ارجاعها را در یک قالب استاندارد آورده شوند.

بارگذاری

- انجام این تمرین به صورت تک نفره است .
- مهلت این تمرین تا پایان روز ۱۹ خرداد ماه است.
- شما قادر نیستید هیچ تمرینی را با بیش از ۷ روز تاخیر بازگذاری کنید. (۷ روز پس از مهلت بارگذاری، سامانه بسته خواهد شد)
- لطفا برای هر تمرین گزارش، فایل کدها و سایر ضمیمات مورد نیاز را با قالب زیر در صفحه درس در سامانه eLearn بارگذاری نمائید .

Example: HW3_810101000_LeCun.zip

ارتباط با تدریسار

- در صورت وجود هر گونه ابهام یا مشکل می‌توانید از طریق رایانامه زیر با دستیاران آموزشی در ارتباط باشید.

❖ fatemehnadi@ut.ac.ir