به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

تمرین پنجم

پرسش ۱	نام دستیار طراح	سیاوش شمس
	رایانامه	siavash.shams@ut.ac.ir
پرسش ۲	نام دستیار طراح	میلاد رئیسی
	رایانامه	miladreisi@ut.ac.ir
	مهلت ارسال پاسخ	۱۴۰۱.۱۰.۱۳

فهرست

قوانین ..... ت

پرسش ۱. آشنایی با مفهوم توجه و پیاده سازی مدل BERT ۱

۱-۱. پیاده‌سازی کدگذار ۱

۱-۲. پیاده‌سازی مدل **BERT** ۲

پرسش ۲ - آشنایی با کاربرد تبدیل کننده‌ها در تصویر ۳

۲-۱. آشنایی با مدل **BEIT** ۳

۲-۲. تقسیم بندی معنایی تصاویر ۳

۲-۳. طبقه بندی تصاویر ۵

۲-۴. پرسش‌ها ۶

شکل‌ها

شکل ۱. بخش کدگذار مدل BERT ۱

شکل ۲. بخش ورودی در مدل BERT. بردار جانمایی (embedding) ورودی برابر است با جمع token embeddings، segment embeddings و position embeddings ۲

شکل ۳. نمونه‌ای از یک تقسیم بندی معنایی ۴

قوانین

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام ***REPORTS_TEMPLATE.docx*** قرار داده شده تهیه نمایید.
- پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین دوم را با شخص B و ... انجام دهید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است**؛ بنابراین، لطفا تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
- **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند**؛ بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- **در صورت مشاهدهٔ تقلب امتیاز تمامی افراد شرکت‌کننده در آن، ۱۰۰- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- **استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست.**
- به ازای هر روز تاخیر ۲ درصد از نمره شما کسر خواهد شد.
- لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

HW[Number] _[Lastname]_[StudentNumber]_[Lastname]_[StudentNumber].zip

(مثال: HW1_Ahmadi_810199101_Bagheri_810199102.zip)

- برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

پرسش ۱. آشنایی با مفهوم توجه[1] و پیاده سازی مدل BERT

در این پرسش قصد داریم تا با پیاده سازی قدم به قدم مدل BERT با نحوه کار آن آشنا شویم. از این رو شما باید کدهای پرونده‌ی **transformer.ipynb** که به پیوست قرار گرفته را با توجه به آموخته‌های خود و مقالاتی که در اختیار شما قرار داده شده کامل کنید و به پرسش‌های زیر پاسخ دهید.

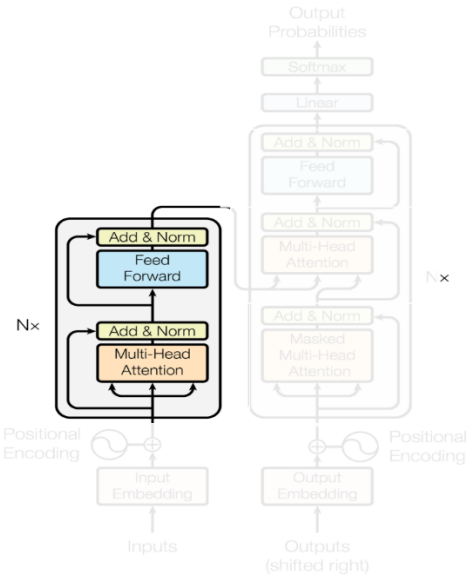
۱-۱. پیاده‌سازی کدگذار[2]

همان‌طور که در درس آشنا شدید بخش کدگذار یک بخش کلیدی در یک تبدیل کننده[3] است. در این بخش با توجه به شکل ۱ قصد داریم تا بخش‌های یک کدگذار را پیاده سازی کنیم. برای این منظور کد مربوط به بخش کدگذار در فایل داده شده را تکمیل کنید.

پرسش‌ها:

۱. توضیح مختصری در مورد مفهوم توجه دهید.

۲. چرا در تبدیل کننده از Multi-head attention به جای Single-head attention استفاده می‌شود؟



شکل ۱. بخش کدگذار مدل **BERT**

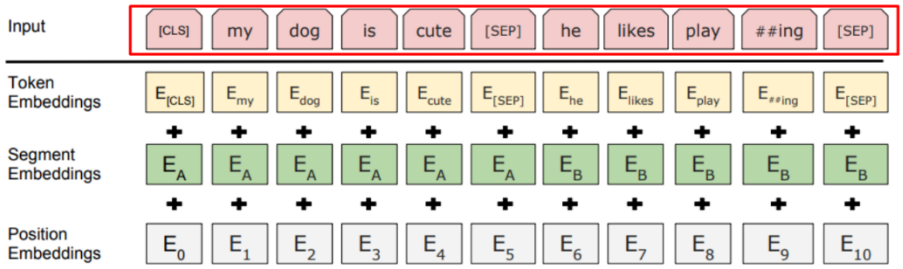
[1] Attention
[2] Encoder
[3] Transformer

۱-۲. پیاده‌سازی مدل BERT

در قسمت قبلی لایه کدگذار را پیاده‌سازی کردید. برای تکمیل پیاده‌سازی مدل BERT مطابق شکل ۲ به دو لایه دیگر نیاز داریم: لایه token embedding و لایه position embedding. (در این تمرین نیازی به پیاده‌سازی بخش segment embeddings ندارید.) مانند قسمت قبل کدهای بخش BERT را در فایل داده شده تکمیل کنید.

۱. در مورد segment embeddings در BERT مختصر توضیح دهید.

۲. پس از تکمیل تمام مراحل و آموزش شبکه روی دادگان داده شده. یک جمله در موردنظر خود در رابطه با یک فیلم (مثلا: "I liked this movie") را به مدل بدهید و خروجی آن را گزارش کنید.



شکل ۲. بخش ورودی در مدل **BERT**. بردار جانمایی (**embedding**) ورودی برابر است با جمع **token embeddings**، **segment embeddings** و **position embeddings**

.

پرسش ۲ – آشنایی با کاربرد تبدیل کننده‌ها[۱] در تصویر

در این بخش ابتدا با یک تبدیل کننده که به تازگی و در سال میلادی جاری، ۲۰۲۲ معرفی شده آشنا خواهید شد، سپس دو تمرین ساده با آن انجام خواهید داد. در نهایت و در بخش آخر به تعدادی پرسش تشریحی و درست-نادرست پاسخ خواهید داد.

۲-۱. آشنایی با مدل BEiT

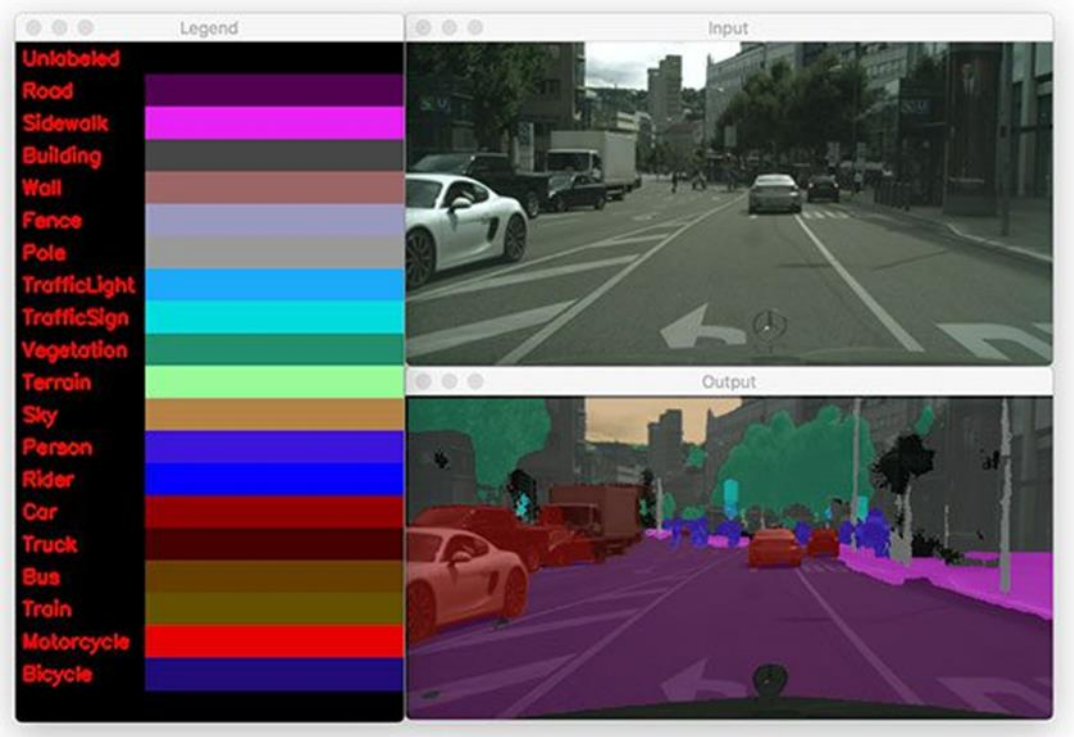
مدل **BEiT** یکی از مدل‌های تبدیل کننده برای تصاویر است که با بهره‌گیری از مدل **BERT** ساخته شده است. مدل **BEiT** به شیوه‌ی خود نظارتی[۲] پیش آموزش[۳] یافته است؛ به این شکل که تصاویر به زیربخش‌های کوچک‌تری نسبت به تصویر اولیه تقسیم شده، سپس چندین بخش آن مخدوش شده، سپس به مدل داده می‌شوند. مدل باید سعی کند تا تصویر اصلی را بازسازی کند. پس از اتمام پیش آموزش، مدل به چند روش دیگر نیز آموزش می‌بیند و در نهایت برای استفاده در کارهای متفاوت مربوط به پردازش تصاویر آماده می‌شود. دو کار متداول در زمینه‌ی پردازش تصویر، طبقه بندی[۴] و تقسیم بندی معنایی[۵] تصاویر با تبدیل کننده‌ها است که در ادامه با آن‌ها آشنا خواهید شد.

۲-۲. تقسیم بندی معنایی تصاویر

تقسیم بندی معنایی، به تقسیم بندی تصویر به زیربخش‌هایی از اجزای مختلفی که در تصویر دیده می‌شوند گویند. شکل ۳ نمونه‌ای از تقسیم بندی معنایی را نشان می‌دهد. در این شکل، تصویر بالا، تصویر اصلی و تصویر پایین، تصویر تقسیم بندی معنایی شده‌ی تصویر اصلی است. در سمت چپ شکل نیز راهنمایی برای تمیز دادن اجزای مختلف تقسیم بندی شده انجام شده قرار داده شده است.

---

۱ Transformers
۲ Self-supervised
۳ Pre-train
۴ Classification
۵ Semantic segmentation



شکل ۳. نمونه‌ای از یک تقسیم بندی معنایی

در این بخش شما باید:

- ابتدا مدل BEIT را با استفاده از مجموعه دادگان تصویری scene_parse_150 بازآموزش دهید.
- سپس ۳ نمونه تصویر از همین مجموعه دادگان به دلخواه انتخاب کنید و اصل تصویر و تقسیم بندی معنایی آن که در مجموعه دادگان داده شده را رسم کنید.
- پس از آن همین تصاویر را با استفاده از مدل بازآموزش یافته‌ای که بدست آوردید تقسیم بندی معنایی کنید و آن را رسم کنید.

بنابراین در مجموع باید ۹ تصویر در گزارش خود بیاورید:

۱. اصل تصاویر دلخواه انتخابیتان

۲. تقسیم بندی معنایی تصاویر که در مجموعه دادگان به صورت پیش فرض داده شده

۳. تقسیم بندی معنایی تصاویر که با مدل بازآموزش داده شده‌ی خود به دست آوردید.

دسترسی به مقاله‌ی BEIT از پیوند زیر ممکن است:

https://arxiv.org/pdf/2106.08254v2.pdf

برای آشنایی بیشتر با مدل BEiT می‌توانید از پیوند زیر استفاده نمایید:

https://huggingface.co/docs/transformers/model_doc/beit

مجموعه دادگان تصویری scene_parse_150 را نیز می‌توانید از طریق پیوند زیر دریافت کنید:

https://huggingface.co/datasets/scene_parse_150

و یا می‌توانید با استفاده از تکه کد زیر آن را در google Colab دریافت کنید:

```
!pip install datasets
from datasets import load_dataset
DS = load_dataset("scene_parse_150")
```

۲-۳. طبقه بندی تصاویر

در این بخش با کاربرد تبدیل کننده‌ها در طبقه بندی تصاویر آشنا خواهید شد. بدین منظور یک طبقه بند ساخته شده با یک تبدیل کننده را با یک طبقه بند ساده مقایسه خواهید کرد. بنابراین شما باید:

- ابتدا باید یک طبقه بند ساده با یک شبکه‌ی **MLP** بسازید.
- سپس یک طبقه بند دیگر نیز با استفاده از مدل **BEiT** آموزش دهید. برای این کار روی مدل **BEiT** یک **MLP** ساده (متشکل از یک یا چند لایه، به دلخواه) قرار دهید.
- در نهایت این طبقه بندها را با هم مقایسه کنید. بدین منظور نمودار دقت و خطا به ازای هر دور را برای داده‌های آموزش و ارزیابی برای هر دو طبقه بند رسم نمایید و ماتریس آشفتگی و گزارش طبقه بندی ها را برای داده‌های ارزیابی ارائه دهید.

در ساخت هر دو طبقه بند از مجموعه دادگان **CIFAR-10** استفاده کنید. این مجموعه دادگان را می‌توانید از پیوند زیر دریافت کنید:

https://www.cs.toronto.edu/~kriz/cifar.html

و یا می‌توانید با استفاده از کتاب‌خانه‌ی **Keras** و تکه کد زیر آن را دریافت کنید:

```
from keras.datasets import cifar10
(X_train, y_train), (X_test, y_test) = cifar10.load_data()
```

تعداد لایه‌ها و دیگر مشخصه‌های[۱] آموزش بر عهده‌ی خودتان است و به دلخواه انتخاب کنید.

هدف آشنایی با عمل طبقه بندی با یک تبدیل کننده است؛ بنابراین نیازی به بهینه سازی عملکرد طبقه بندها نیست و سعی کنید در کوتاه ترین زمان ممکن آموزش آن‌ها را انجام دهید.

نیازی به پیش پردازش خاصی برای تصاویر نیست و فقط در صورت نیاز پیش پردازش‌های لازم برای مدل **BEIT** را انجام دهید.

۲-۴. پرسش‌ها

در این بخش به تعدادی پرسش تشریحی و پرسش درست-نادرست پاسخ خواهید داد. تمامی پرسش‌ها از جلسات ضبط شده‌ی درس ایده برداری شده است.

به پرسش‌های زیر پاسخ دهید: (نهایت ۳ خط)

۱. در شبکه‌های CNN در کدام بخش مفهومی مانند مفهوم توجه اتفاق می‌افتد؟

۲. در یک شبکه‌ی عصبی، در ارتباط یک لایه با لایه‌ی بعد، چه تفاوتی میان یک شبکه‌ی convolution با شبکه‌ی توجه همگانی[۲] و شبکه‌ی توجه محلی[۳] وجود دارد؟

درست یا نادرستی جملات زیر را تعیین کنید:

۱. در بخشی از لایه‌های تبدیل کننده‌ی Vanilla از شبکه‌ی LSTM استفاده شده است.

---

[۱] Parameters

[۲] Global attention

[۳] Local attention

۲. یک تبدیل کننده از چند بلوک رمزگذار[۱] و چند بلوک رمزگشا[۲] تشکیل شده است.

۳. multi head attention از یک بخش توجه[۳] و چند لایه‌ی تمام متصل[۴] موازی تشکیل شده است.

۴. وجود Positional Encoding در ساختار یک تبدیل کننده حیاتی است و بدون آن شبکه از کار می‌افتد.

---

۱ Encoder

۲ Decoder

۳ Attention

۴ Fully connected