

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

**درس شبکه‌های عصبی و یادگیری عمیق**  
**تمرین امتیازی**

رایانامه	نام دستیار طراح	
fatemehmirzadeh99@gmail.com	فاطمه میرزاده	پرسش ۱
Baledi.danial@gmail.com	دانیال بالدی	پرسش ۲
m.nourbakhsh75@gmail.com	مهرداد نوربخش	پرسش ۳
۱۴۰۲.۱۰.۲۱		مهلت ارسال پاسخ

قوانین.....	۱
پرسش ۱. آشنایی با شبکه عصبی کانولوشن برگشتی.....	۱
۱-۱. توضیحات مدل.....	۱
۲-۱. مجموعه دادگان و پیش‌پردازش آنها.....	۱
۳-۱. آموزش مدل.....	۲
۴-۱. ارزیابی و تحلیل نتایج.....	۲
پرسش ۲ - LoRA.....	۳
۱-۲. LoRA چگونه عمل می‌کند؟.....	۳
۲-۲. قرار است چه کاری را روی چه داده‌هایی انجام دهیم؟.....	۳
۳-۲. و بالاخره کد نویسی: آموزش مدل.....	۴
۴-۲. چرا LoRA؟.....	۴
نحوه‌ی انتخاب دیتاست.....	۵
پرسش ۳ - تشخیص تقلب.....	۶
۱-۳. آشنایی با دیتاست.....	۶
۲-۳. پیاده‌سازی معماری مقاله.....	۶
۳-۳. نمونه‌برداری.....	۷
۴-۳. آموزش مدل با داده‌های جدید.....	۷

## شکل‌ها

شکل ۱. تبدیل جملات به شاخص‌های کلمه ..... ۲

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 ... هم‌گروه باشید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفاً 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.
- **کدها حتماً باید در قالب نوت‌بوک با پسوند ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتماً در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کدها وجود داشته باشد.
- **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت‌کننده در آن، 100- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- **استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست.**

- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر (به ازای هر روز یک درصد تجمعی) وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر 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 شبکه عصبی کانولوشن برگشتی

هدف از این تمرین آشنایی با یک مدل جدید به اسم BRCNN (شبکه عصبی کانولوشن برگشتی) و پیاده‌سازی و بکارگیری آن در زمینه پردازش زبان طبیعی (NLP) است. در این بخش به پیاده‌سازی مقاله‌ای با عنوان زیر خواهید پرداخت که به پیوست نیز فرستاده شده است:

Bidirectional Recurrent Convolutional Neural Network for Relation Classification

### ۱-۱. توضیحات مدل

(۱۰ نمره)

مقاله را با دقت مطالعه کرده و راجب مدل بکار رفته در مقاله توضیحات مختصری داده و نحوه عملکرد آن را شرح دهید.

### ۲-۱. مجموعه دادگان و پیش‌پردازش آنها

(۴۵ نمره)

الف. جهت بارگزاری و بارگیری دیتاست مورد استفاده می‌توانید از نمونه کد قرار گرفته به نام Bidirectional\_Dataset که به پیوست فرستاده شده است استفاده نمایید (توجه داشته باشید هم داده‌های آموزش و هم آزمایش را آپلود کنید). پس از در نظر گرفتن دادگان، تعداد جملات آموزشی (training)، آزمایشی (test)، تعداد کلاس‌ها و همچنین تعداد نمونه در هر کلاس را بدست آورده و ترسیم نمایید. (۱۰ نمره)

ب. پیش‌پردازش داده‌ها: باید جملات خود را به شاخص‌های کلمه همچون شکل ۱-۱ تبدیل کنید. هر کلمه دارای یک شاخص منحصر به فرد در واژگان ساخته شده از مجموعه آموزش است. برای اطمینان از اینکه همه جملات شکل یکسانی دارند، لیست‌ها تا حداکثر طول با صفر پر می‌شوند. (۱۵ نمره)

"The NYT is a famous newspaper."  
 1 4 3 7 42 15 0 0 0 0

"The newspaper shares the life of famous people."  
 1 15 18 1 9 22 42 1 0 0

شکل ۱. تبدیل جملات به شاخص‌های کلمه

می‌توانید اینکار را با استفاده از Keras Tokenizer انجام دهید.

ج. قبل از اینکه به سراغ مدل واقعی برویم، باید برجسب‌های داده‌ها را به شکلی قابل فهم برای کامپیوتر تبدیل کنیم. این کار انواع مختلفی دارد. در این جا برجسب‌های ما باید به کدگذاری یک‌طرفه تبدیل شوند. به عنوان مثال، ۰ [۰، ۰، ۰، ...، ۰، ۱، ۰، ۰، ۰، ۰] تبدیل به ۳ و [۰، ۰، ۰، ...، ۰، ۰، ۰، ۰، ۰، ۰] شود. (۲۰ نمره)

### ۳-۱. آموزش مدل

(۲۵ نمره)

مدل شرح داده شده در مقاله را پیاده سازی کرده و سپس برای ۲۰ دور آن را آموزش دهید.

### ۴-۱. ارزیابی و تحلیل نتایج

(۲۰ نمره)

همانطور که قبلاً گفته شد، دقت به تنهایی واقعاً عملکرد مدل را منعکس نمی‌کند، بنابراین استفاده از تابع ارزیابی کافی نیست. پس، از تابع پیش بینی استفاده کنید تا احتمال هر کلاس را خروجی را بدست آورید. ماتریس درهم‌ریختگی و دقت، فراخوانی و امتیاز f1 (f1-score، recall، precision) را برای هر کلاس ارائه دهید.

قبلا با Transfer learning و برخی روش‌های fine tuning آشنا شدیم. مثلا fine tune کردن کل پارامترهای مدل یا لایه‌هایی خاص از مدل از این روش‌ها هستند. در این تمرین می‌خواهیم روش متفاوت‌تری را بررسی کنیم. این روش low rank adaptation of large language models یا به اختصار LoRA نام دارد. مقاله‌ی مرتبط با آن در پیوست آمده است و از این [لینک](#) نیز قابل دریافت است. برای حل سوال خواندن و فهمیدن تمام جزئیات مقاله لازم نیست هر چند خواندن تمام آن توصیه می‌گردد.

### ۲-۱. LoRA چگونه عمل می‌کند؟

(۲۰ نمره)

ابتدا دو روش ذکر شده در مقدمه‌ی سوال، یعنی fine tune کل پارامترهای مدل و fine tune کردن یک یا چند لایه از لایه‌های مدل را مختصرا توضیح دهید. پس از توضیح دو روش ذکر شده، بیان کنید که روش LoRA برای fine tune کردن چگونه کار می‌کند.

### ۲-۲. قرار است چه کاری را روی چه داده‌هایی انجام دهیم؟

(۲۰ نمره)

در این سوال قرار است از مدل زبانی بزرگ<sup>۱</sup> RoBERTa<sup>۲</sup> نسخه‌ی Large استفاده کنید. این مدل را می‌توان به صورت آماده و با دستور زیر دریافت کرد (می‌توانید از هر روش دیگری نیز استفاده کنید ولی در این صورت از سالم بودن مدل دریافت شده اطمینان حاصل نمایید).

```
from transformers import AutoTokenizer, AutoModel

tokenizer = AutoTokenizer.from_pretrained("roberta-large")
model = AutoModel.from_pretrained("roberta-large")
```

<sup>۱</sup> Large Language Model (LLM)

<sup>۲</sup> RoBERTa: A Robustly Optimized BERT Pretraining Approach



حال با توجه به توضیحات انتهایی سوال، در رابطه با دیتاست منتصب شده به گروه‌تان توضیح مختصری ارائه دهید. نیازی به بررسی جزئیات دیتاست نیست و توضیحات کلی در رابطه با هدف دیتاست و نوع داده‌های آن کفایت می‌کند. پس از بررسی دیتاست بیان کنید که قرار است از چه روشی برای آموزش مدل جهت وظیفه‌ی مدنظر دیتاست استفاده کنید (یعنی توضیح دهید ورودی مدل چیست؟ چگونه از خروجی مدل استفاده می‌کنید؟ آیا نیاز به اعمال تغییر در مدل است؟ اگر بله چه تغییری؟ LoRA چگونه باید روی مدل اعمال گردد (با توجه به مقاله)؟ و توضیحاتی از این دست. در این قسمت نیازی به کد نویسی نیست) [بررسی دیتاست ۵ نمره – باقی سوال ۱۵ نمره].

## ۲-۳. و بالاخره کد نویسی: آموزش مدل

(۴۰ نمره)

یک بار با fine tune تمامی پارامترهای مدل و یک بار دیگر با اعمال LoRA و freeze کردن تمام پارامترهای غیر LoRA مدل را آموزش دهید سپس معماری دو مدل، زمان صرف شده جهت آموزش دو مدل، تعداد پارامترهایی که آموزش می‌بینند و دقت دو روش را مقایسه کنید (می‌توانید مقایسه را در بخش بعدی سوال نیز انجام دهید). هاپرپارامتر<sup>۱</sup>های آموزش مدل در مقاله آمده‌اند ولی در صورت داشتن محدودیت پردازشی، به مقدار ایپاک کمتر یا روی داده‌های کم‌تر مدل را آموزش دهید.

می‌توانید از توابع آماده برای پیاده سازی و اعمال LoRA استفاده کنید. همچنین برای آموزش مدل می‌توانید از توابع trainer آماده استفاده کنید (ولی استفاده از pipeline های huggingface مجاز نیست) [آموزش مدل و گزارش نتایج ۳۰ نمره – مقایسه دو مدل ۱۰ نمره].

## ۲-۴. چرا LoRA؟

(۲۰ نمره)

روش LoRA را با دو روشی که در بخش اول توضیح دادید مقایسه کنید. در مقایسه این مورد را نیز ذکر کنید که اگر بخواهیم بدون fine tune مجدد پارامترهای اصلی یک مدل (مثلا RoBERTa) جهت آموزش و inference چند وظیفه استفاده کنیم (مثلا از همان یک مدل هم برای وظیفه‌ی تحلیل احساسات استفاده

---

<sup>1</sup> Hyper parameter

کنیم و هم برای وظیفه‌ی خلاصه سازی متن)، در هر یک از روش‌ها این کار چگونه انجام می‌گردد؟ اصلاً چنین چیزی امکان پذیر است؟

(توصیه) می‌توانید از تجربه‌ی خودتان از بخش قبلی و توضیحات مقاله برای پاسخ این بخش استفاده کنید

## نحوه‌ی انتخاب دیتاست

اگر شماره دانشجویی فرد باشد از دیتاست [Multi NLI](#)<sup>۱</sup> و اگر زوج باشد از دیتاست [QQP](#)<sup>۲</sup> استفاده کنید. ملاک شماره دانشجویی کوچکتر در گروه است. مثلاً در گروه با دو شماره دانشجویی: ۸۱۰۱۰۱۰۱۱ و ۸۱۰۱۰۱۱۲۲ شماره دانشجویی اول (۸۱۰۱۰۱۰۱۱) ملاک است و دیتاست Multi NLI به این گروه منتصب می‌گردد.

(توجه) می‌توانید دیتاست‌ها را از منابع دیگر نیز دریافت کنید. در این صورت از صحت آن‌ها اطمینان یابید

---

<sup>۱</sup> <https://cims.nyu.edu/~sbowman/multinli/>

<sup>۲</sup> <https://quoradata.quora.com/First-Quora-Dataset-Release-Question-Pairs>

## پرسش ۳ - تشخیص تقلب

روزانه میلیون‌ها تراکنش بانکی از طریق کارت‌های اعتباری و سرویس‌های آنلاین در دنیا انجام می‌شود. یکی از مشکلاتی که در تراکنش‌های بانکی وجود دارد، رخ دادن تقلب یا fraud است. تراکنش‌های نامعتبر می‌توانند باعث زیان مالی افراد شوند. بنابراین لازم است تا مکانیزم‌هایی وجود داشته باشد تا اینگونه تقلب‌ها را تشخیص دهند و مانع ضرر افراد شوند. در این تمرین قصد داریم تا با استفاده از شبکه‌ی عصبی، یک سیستم تشخیص تقلب را پیاده‌سازی و بررسی کنیم. مقاله‌ای که قصد پیاده‌سازی آن را داریم در این [لینک](#) آورده شده است.

### ۱-۳. آشنایی با دیتاست

(۱۰ نمره)

دیتاستی که قرار است از آن استفاده کنیم در این [لینک](#) آورده شده است. این دیتاست شامل داده‌های مربوط به تراکنش‌های بانکی است که توسط دارندگان کارت اعتباری انجام شده است. دیتاست رو بارگذاری کرده و به سوالات زیر جواب دهید.

۱. نمودار هیستوگرام کلاس‌های داده را رسم کرده و آن را تحلیل کنید.

۲. اگر مدل را با این داده‌ها آموزش دهیم چه مشکلی به وجود می‌آید؟ توضیح دهید.

### ۲-۳. پیاده‌سازی معماری مقاله

(۳۵ نمره)

معماری معرفی شده در مقاله را به طور مختصر توضیح دهید. سپس آن را پیاده‌سازی کنید و با انجام پیش‌پردازش‌های لازم برای داده‌های قسمت قبل، مدل را آموزش دهید و سپس موارد زیر را گزارش کنید. (هر پارامتری را که برای یادگیری استفاده کرده‌اید گزارش کنید).

۱. نمودار loss و accuracy را برای داده‌های آموزش و ارزیابی رسم کنید و آن را تحلیل کنید. توضیح دهید که آیا مدل overfit شده است یا نه.

۲. پیش‌بینی مدل برای داده‌های تست را بدست آورید و نمودار heatmap را برای confusion matrix رسم کنید. همچنین معیارهای precision, recall و f1-score را گزارش کنید.
۳. توضیح دهید که آیا accuracy می‌تواند به تنهایی عملکرد مدل را به خوبی نمایش دهد؟

### ۳-۳. نمونه‌برداری

(۳۵ نمره)

- یکی از روش‌هایی که برای متوازن کردن داده‌های نامتوازن استفاده می‌شود، روش نمونه‌برداری است. در مقاله از روش Adaptive Synthetic Sampling استفاده شده است که در این بخش به آن می‌پردازیم.
۱. این روش را توضیح دهید و مزایا و معایب آن را نسبت به روش‌های دیگر نمونه‌برداری بیان کنید.
  ۲. این روش را پیاده‌سازی کنید و به طور مختصر روش خود را توضیح دهید. توجه کنید که در این قسمت نمی‌توانید از کتابخانه‌های آماده استفاده کنید.
  ۳. توضیح دهید که آیا نمونه‌برداری باید قبل از تقسیم کردن داده‌ها به داده‌های آموزش و تست انجام شود یا بعد از آن؟
  ۴. با استفاده از روش پیاده‌سازی شده، نمونه‌برداری را برای داده‌های دیتاست انجام دهید و هیستوگرام کلاس‌ها را برای داده‌های جدید رسم کنید.

### ۳-۴. آموزش مدل با داده‌های جدید

(۲۰ نمره)

- پس از آن که نمونه‌برداری را انجام دادید، مدل را مجدداً آموزش دهید و موارد خواسته شده در بخش اول را گزارش کنید. نتایج بدست آمده از دو مدل را با یکدیگر مقایسه و تحلیل کنید.
- (توجه)** در هیچ‌کدام از بخش‌های سوال، رسیدن به دقت مقاله مد نظر نیست.