

به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



یادگیری عمیق پیشرفته

پروژه اول

اسفند ۹۹

## فهرست سوالات

- ۳..... Probabilistic Approach to Embedding – سوال ۱ (الزامی)
- ۳..... بخش اول
- ۳..... بخش دوم
- ۵..... Transfer Learning using EfficientNet-B0 – سوال ۲ (انتخابی)
- ۵..... مقدمه
- ۵..... بخش اول
- ۶..... بخش دوم
- ۶..... بخش سوم
- ۷..... Sentiment Analysis with Pretrained BERT – سوال ۳ (انتخابی)
- ۷..... مقدمه
- ۷..... بخش اول
- ۷..... بخش دوم
- ۸..... بخش سوم
- ۹..... Embedding توسط شما – سوال ۴ (الزامی) تعریف پروژه در کاربرد

### توجه:

پاسخ گویی به سوال های شماره ۱ و ۴ الزامی است.

از بین مسائل شماره ۲ و ۳ هم یکی را باید انتخاب کنید. البته انجام هر دو سوال امتیاز ویژه ای دارد، به این صورت که اگر هر دو سوال را انجام دهید، بیشینه امتیاز کسب شده از میان آنها برای شما منظور خواهد شد.

## الزامی) سوال ۱ – Probabilistic Approach to Embedding

### بخش اول

فرض کنید عاملی داریم که با استفاده از داده‌های ورودی یادگیری را انجام می‌دهد و هدف عامل نیز کاهش خطای مربعات (MSE) است. از دیدگاه آماری، وقتی مقدار واقعی خروجی از مقدار تخمین آن کم می‌شود، به پارامتری به نام باقی مانده (residual)، می‌رسیم که خود متغیری تصادفی است.

الف) رابطه کلی بین باقی مانده و تخمین‌مان از مقدار خروجی را پیدا کنید. پاسخ خود را توجیه کنید و برای توجیه تا حد امکان از روابط ریاضی بهره بگیرید.

ب) با توجه به اینکه تابع خطای ما برابر توان دوم مقدار باقی مانده است، طبق رابطه :

$$e = y - \bar{y}$$

$$E[e^2] = \text{var}(e) + b^2[e]$$

می‌توان نتیجه گرفت که این مقدار خطا برابر با مجموع توان دوم بایاس و واریانس باقی‌مانده است. فرض کنید که  $2n$  داده آموزش داریم و از  $h$  داده تست برای ارزیابی کارایی عامل بهره می‌گیریم. یک بار عامل را با  $2n$  داده آموزش می‌دهیم و یک بار دیگر با  $n$  داده آموزش می‌دهیم. سپس با عامل آموزش دیده، خروجی را برای  $h$  داده تست حساب می‌کنیم. مقدار بایاس را در دو حالت بررسی کنید. بررسی خود را در حالات مختلف با روابط ریاضی توجیه کنید.

### بخش دوم

یکی از کاربردهای embedding، استخراج اطلاعات از عکس و پرسش و سپس یافتن پاسخ مناسب است (visual question answering). در کاربرد مذکور، از عاملی برای embedding عکس و پرسش استفاده می‌شود، بردار embedding عکس و پرسش به بردار embedding پاسخ map می‌شود و سپس با عاملی طبقه بند، بردار embedding ایجاد شده در دسته یکی از پاسخ‌ها طبقه بندی می‌شود که پاسخی نهایی است. مقاله‌های زیادی در این خصوص موجود هستند که می‌توانید با مراجعه به آن‌ها ساز و کار و روابط را به صورت کامل یاد بگیرید.

الف) کار با چنین عامل‌های هوشمندی را توضیح دهید و روابط و ساختار مربوط به آنان را بیابید و آن‌ها را ضمیمه گزارش کنید.

ب) اگر از دیدگاه آماری به قضیه نگاه کنید، متوجه خواهید شد که به نحوی می‌خواهیم با استفاده از توزیعی از دادگان ورودی که شامل عکس و پرسش‌ها است، توزیعی از خروجی که دادگان پاسخ هستند را

بیابیم. یک مدل آماری بسازید به نحوی که به خوبی بتواند منطق پیش بینی خروجی را به خواننده برساند. (به عنوان مثال شما میتوانید از مدل های آماری Bayesian استفاده کنید تا با استفاده از توزیع حال حاضر بردار embedding ورودی، توزیعی متناسب بردار embedding خروجی را پیش بینی کنید؛ یا اینکه از طبقه بند نزدیک ترین همسایگان (KNN) استفاده کنید تا نزدیک ترین بردارهای embedding خروجی را بیابید. دقت کنید که این موارد تنها راه حل های سوال نیست).

ج) به سورس کد کلاس لایه [Embedding](#) مراجعه کرده و ضمن توضیح کامل این کلاس در گزارش کار خود، با روش ارث بری در برنامه نویسی شیء گرا، یک method به این کلاس اضافه کنید که با استفاده از آن بتوان میزان شباهت به بردارهای embedding دیگر را به کاربر گزارش داد. این معیار شباهت میتواند معیار فاصله اقلیدسی یا فاصله کسینوسی باشد.

د) با تحقیق در منابع، مجموعه دادگانی که شامل عکس، پرسش و پاسخ هستند را بیابید و توضیحی مفصل در خصوص دادگان در گزارش کار خود بیاورید. سپس کلاسی به نام revised embedding بسازید که در آن از مدل آماری پیشنهادی شما در بند ب استفاده شده باشد. دقت کنید که کلاس ذکر شده باید شامل الگوریتم وصف شده شما در مرحله دوم باشد و در صورتی که به سورس کد embedding اضافه شود، قابل فراخوانی باشد و کاربر بتواند از آن استفاده کند. در نهایت، با استفاده از کلاس ساخته شده، عملکرد مدل پیشنهادی خود را بررسی کنید و گزارش دهید.

## انتخابی) سوال ۲ – Transfer Learning using EfficientNet-B0

### مقدمه

مدل EfficientNet-B0 یک شبکه عصبی Convolutional و از خانواده هفتگانه مدل‌های [EfficientNet](#) است. این مدل‌ها با تمرکز بر کاربرد دسته‌بندی تصاویر طراحی شده‌اند و هر کدام دقت مختلفی در دسته‌بندی تصاویر دارند؛ به طور عمده تفاوت آنها در تعداد پارامترهای هر کدام است. محققین زیادی این مدل‌ها را روی مجموعه‌های داده متفاوتی از قبل آموزش داده‌اند و وزن‌های حاصل از این آموزش‌ها روی اینترنت در دسترس است.

حالا ما می‌خواهیم از یک مدل EfficientNet-B0 را که با وزن‌های مجموعه دادگان عظیم ImageNet آموزش داده شده، در دسته‌بندی تصاویر جدیدی استفاده کنیم. که برای دسته‌بندی تصاویر جدید از embedding موجود در EfficientNet-B0 بهره می‌بریم.

سپس در سه سطح روی این مدل fine-tuning انجام می‌دهیم تا در نهایت دقت بالاتری در دسته‌بندی ۱۹۶ مدل خودرو در مجموعه دادگان [Stanford Cars dataset - car196](#) به دست آید.

### بخش اول

روش کار باید به صورت زیر باشد:

۱. ابتدا مدل را بدون لایه‌های Fully connected انتهایی و تنها با لایه‌های بلوک‌های convolutional بارگذاری کنید. پیشنهاد می‌شود در بارگذاری مدل از وزن‌های [NoisyStudent](#) بهره ببرید که دقت بالاتری به دست می‌دهند ([در باره NoisyStudent](#)).  
۲. یادگیری را برای تمام لایه‌های تعریف شده مذکور غیرفعال کنید (اصطلاحاً وزن لایه‌ها را freeze کنید).

۳. سپس خروجی فیلترهای آخرین لایه convolutional را به یک معماری Classifier مناسب دلخواه متصل کنید. این معماری می‌تواند لایه‌های fully-connected و یا یک مدل KNN باشد که پیاده‌سازی می‌کنید.

۴. مجموعه دادگان [Cars dataset](#) را بارگذاری کنید. با تصاویر این مجموعه، آموزش شبکه را آغاز کنید (این اولین سطح fine-tuning شبکه خواهد بود).

دقت و خطای شبکه را در پیشبینی روی دادگان تست را گزارش کنید.

## بخش دوم

پس از قدم‌های ذکر شده بالا، حالا سعی می‌کنیم برای به دست آوردن دقت بیشتر، در دو حالت زیر شبکه را مجدداً آموزش دهیم (این دو حالت را جداگانه انجام دهید تا تاثیر هر کدام به طور مستقل معلوم شود):

- یکبار کل لایه‌های قبلی را unfreeze کنید و آموزش را تکرار کنید (آنها را trainable کنید).
- و یکبار دیگر فقط لایه‌های convolutional بلوک ششم و هفتم و آخرین لایه (top\_conv) را unfreeze کنید و آموزش را تکرار کنید.

دقت داشته باشید که در هر دوی این حالت‌ها، وزن‌های بخش classifier قابل آموزش هستند. از دیدگاه ویژگی‌های لازم برای به دست آوردن representation خوب و ارتباط آن با ویژگی‌های مدل که در کلاس درس توضیح داده شد، چگونه می‌توان تغییر به دست آمده در دقت شبکه در بخش دوم را تحلیل کرد؟

## بخش سوم

با توجه به افزایش دقتی که در بخش دوم حاصل می‌شود، پیشنهادهای خود را برای بالا بردن بیشتر دقت با استفاده از تکنیک‌های مختلف ارائه کنید. لطفاً توضیح دهید که از هر تکنیک پیشنهادی شما به چه دلیل می‌بایست انتظار بهبود در نتیجه پیش بینی دسته بندی‌های شبکه را داشته باشیم.

## انتخابی) سوال ۳ – Sentiment Analysis with Pretrained BERT

### مقدمه

در این پروژه، هدف ما پیاده سازی Sentiment Analysis یا تشخیص تحلیل احساسات درون جمله‌ها است، در واقع هر جمله در توصیف یک فیلم گفته شده و باید با توجه به جمله، تشخیص دهیم که حس گوینده نسبت به آن فیلم چه بوده است.

در این مسئله انتظار داریم که شما از مدل BERT استفاده کنید. BERT یک مدل مبتنی بر Transformer است که به شکل آموزش دیده شده در کاربردهای مختلف پردازش زبان طبیعی استفاده می‌شود. در واقع این مدل representation مدل زبان را در داخل خود دارد و ما با استفاده از آن می‌توانیم کاربردهای دیگری مانند سیستم‌های سوال و پاسخ یا تحلیل متون را انجام دهیم. برای مطالعه بیشتر درباره این مدل، به [مقاله معرفی BERT](#) مراجعه کنید.

### بخش اول

مجموعه داده‌ای که در این سوال استفاده خواهیم کرد، در [این جا](#) قرار دارد، و هدف ما استفاده شما از یک مدل BERT از پیش آموزش دیده برای استخراج embedding مناسب از جملات و سپس طبقه بندی آن‌ها است.

### بخش دوم

پس از خواندن مجموع داده و tokenize کردن آن برای BERT باید به دو سوال پاسخ دهید:

۱. ورودی را به چه شکلی به شبکه بدهیم، همانطور که میدانیم، BERT دو جمله ورودی می‌گیرد ولی ما اینجا تنها یک جمله داریم. (دلیل این که معمولاً کسی در این رابطه چیزی نمی‌نویسد این است که معمولاً BERT را fine-tune می‌کنند ولی ما نمی‌خواهیم وزن‌های شبکه را تغییر دهیم و تنها به دنبال استخراج embedding هستیم).

۲. چه خروجی‌ای از شبکه بگیریم؟ هرچند که معمولاً خروجی [CLS] را از آن می‌گیرند، ولی این روش الزاماً بهینه نیست (مخصوصاً وقتی fine-tune نمی‌کنیم).

برای هر کدام از دو سوال بالا، چند راه حل پیشنهاد دهید؛ و پس از مقایسه روش‌های مختلف، سعی کنید دقت مدل‌های مختلف را توضیح دهید.

برای سرعت بخشیدن به انجام آزمایش‌های مختلف، می‌توانید از یک SVM برای classification روی ویژگی‌های frozen استفاده کنید.

### **بخش سوم**

در مرحله آخر، قصد داریم با افزایش حجم داده‌ها، دقت را بالاتر ببریم، برای این کار دو روش افزایش داده برای متن را پیاده کنید و دقت‌ها را مقایسه کنید (اجباری وجود ندارد که به دقت بالاتر برسید، صرفاً هدف آشنایی با روش‌های افزایش داده متنی است).



## (الزامی) سوال ۴ – تعریف پروژه در کاربرد Embedding توسط شما

در زمینه embedding موضوعی جدید ارائه دهید و ایده خود را پیاده سازی کنید!

موضوع شما میتواند به صورت ایجاد نوآوری در قسمت تئوری، طراحی ساختار، کاربرد و هدف جدید و غیره باشد. اما موضوع مورد توجه این است که شما باید کار نسبتاً جدیدی ارائه دهید. بنابراین توصیه نمی‌شود که الگوریتم‌های پیاده سازی شده یا طراحی شده را مجدداً به همان شکل استفاده کنید چرا که در این صورت نمره قسمت نوآوری را از دست می‌دهید!

انتظار می‌رود که پروپوزال شما شامل سه بخش زیر باشد.

۱. طرح صورت مسئله: بگویید چه چیزی را می‌خواهید حل کنید.
۲. مدل pre-trained: مدلی که می‌خواهید برای گرفتن embedding از آن بهره ببرید، کدام است.
۳. مجموعه دادگان: مجموعه داده ای که از ویژگی‌های embedded آن می‌خواهید استفاده کنید را معرفی کنید. همچنین مجموعه داده ای که می‌خواهید روی آن fine-tune انجام دهید را معرفی کنید (یا تعریف ورودی ای که می‌خواهید بر اساس آن تولید داده انجام دهید).

تاریخ ارسال پروپوزال: ۱۳ فروردین ۱۴۰۰

تاریخ ارسال گزارش کامل: همراه با گزارش کل پروژه

## نکات:

- مهلت تحویل این پروژه ۳۱ فروردین ۱۴۰۰ است.
- گزارش را در قالب تهیه شده که روی صفحه درس در Elearn بارگذاری شده، بنویسید.
- گزارش شما در فرآیند تصحیح از اهمیت ویژهای برخوردار است. لطفاً تمامی نکات و فرضیهایی که برای پیاده سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
- الزامی به ارائه توضیح جزئیات خط به خط کد در گزارش نیست. اما باید نتایج بدست آمده را گزارش و تحلیل کنید.
- هرگونه نتیجه و یا تحلیلی که در شرح سوال از شما خواسته شده است را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر میشود.
- برای انجام پروژه‌ها، تنها زبان برنامه نویسی مجاز Python است.
- در صورت مشاهده تقلب، امتیاز تمامی افراد شرکت کننده در آن ۱۰۰- لحاظ میشود.
- در بخش‌های محدودی از کد و فقط به عنوان راهنمایی برای پیاده سازی، مجاز خواهید بود که از کدهای آماده استفاده کنید.
- پاسخ‌گویی به سوال‌های شماره ۱ و ۴ الزامی است.
- از بین مسائل شماره ۲ و ۳ هم یکی را باید انتخاب کنید. انجام هر دو سوال امتیاز ویژه‌ای دارد، به این صورت که اگر هر دو سوال را انجام دهید، بیشینه امتیاز کسب شده از میان آنها برای شما منظور خواهد شد.
- نحوه محاسبه تاخیر به این شکل است: پس از اتمام مهلت اصلی ارسال، به مدت سه روز (تا ۳ اردیبهشت) بارگذاری با تاخیر ممکن است اما به ازای هر روز ۲۰ درصد از نمره کسر خواهد شد؛ در نهایت و پس از بازه تاخیر، ارسال ممکن نیست و نمره تکلیف صفر خواهد شد.
- لطفاً گزارش، فایل کدها و سایر ضمایم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمایید.

PROJECT#\_[Lastname]\_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل میتوانید از طریق رایانامه‌های زیر با دستیاران آموزشی مربوطه آقایان حمیدرضا هاشم‌پور (سوال اول)، حامد آهنگری (سوال دوم و چهارم) و سپهر سامنی (سوال سوم) در تماس باشید:

[hamidreza.hashemp@ut.ac.ir](mailto:hamidreza.hashemp@ut.ac.ir)

[h.ahangari@ut.ac.ir](mailto:h.ahangari@ut.ac.ir)

[sepehr.sameni@gmail.com](mailto:sepehr.sameni@gmail.com)