



به نام خدا  
دانشگاه تهران  
دانشکده مهندسی برق و کامپیوتر



## درس شبکه‌های عصبی و یادگیری عمیق

### تمرین پنجم

نام و نام خانوادگی	امیرحسین پورداود
شماره دانشجویی	۸۱۰۱۰۱۱۲۰
تاریخ ارسال گزارش	۱۴۰۱.۱۰.۱۶

## فهرست

پاسخ ۱. آشنایی با مفهوم توجه و پیاده سازی مدل BERT.....	۴
۱-۱. پیاده سازی کدگذار (Encoder).....	۴
۱-۱-۱. توضیح مختصر درباره توجه (Attention).....	۴
۱-۱-۲. دلیل استفاده از Multi-head attention به جای Single-head attention.....	۵
۲-۱. پیاده سازی مدل BERT.....	۶
۱-۲-۱. توضیح درمورد segment embedding در BERT.....	۶
۲-۲-۱. نمونه خروجی در رابطه با یک فیلم.....	۷

## شکل‌ها

شکل ۱ - نمونه خروجی شبکه BERT و مقدار توجه بر روی دو کلمه مختلف.....۷

## جدول‌ها

جدول ۱ - جدول نمونه segment embedding ..... ۶

## پاسخ ۱. آشنایی با مفهوم توجه و پیاده سازی مدل BERT

### ۱-۱. پیاده سازی کدگذار (Encoder)

کد مربوط به بخش کدگذار در فایل داده شده، تکمیل شده است و در پیوست ارسال شده است.

#### ۱-۱-۱. توضیح مختصر درباره توجه (Attention)

در روانشناسی، توجه فرآیند شناختی تمرکز انتخابی بر یک یا چند چیز در حالی که دیگران را نادیده می گیرد. در یادگیری عمیق، مکانیسم توجه همچنین تلاشی برای اجرای عمل تمرکز انتخابی بر روی چند چیز مرتبط است.

با یک مثال توضیح می دهیم:

فرض کنید در حال دیدن یک عکس دسته جمعی از اولین مدرسه خود هستید. به طور معمول، گروهی از بچه ها در چند ردیف می نشینند و معلم جایی در بین آنها می نشیند. حالا اگر کسی این سوال را بپرسد که «چند نفر هستند؟» چگونه به آن پاسخ می دهید؟ به سادگی با شمارش سرها، نیازی نیست چیزهای دیگری را در عکس در نظر بگیرید. حال، اگر کسی سوال دیگری بپرسد، «معلم در عکس کیست؟»، مغز شما دقیقاً می داند چه باید بکند. به سادگی شروع به جستجوی ویژگی های یک بزرگسال در عکس می کند. بقیه ویژگی ها به سادگی نادیده گرفته می شوند. این همان «توجه» است که مغز ما در اجرای آن بسیار ماهر است.

توجه به طور کلی به توانایی تمرکز بر یک چیز و نادیده گرفتن چیزهای دیگر که در آن زمان بی ربط به نظر می رسند اشاره دارد. در یادگیری ماشینی، این مفهوم با آموزش مدل برای تمرکز بر بخش های خاصی از داده های ورودی و نادیده گرفتن سایر بخش ها برای حل بهتر کار در دست اعمال می شود.

برای مثال، در کارهایی مانند ترجمه ماشینی، داده های ورودی دنباله ای از متن هستند. وقتی ما انسان ها یک متن را می خوانیم، طبیعی به نظر می رسد که به برخی از قسمت ها بیشتر از بقیه توجه کنیم. معمولاً قسمتی از جمله کی، کی و کجاست که توجه ما را به خود جلب می کند.

از آنجایی که این مهارتی است که ما از بدو تولد ایجاد می کنیم، اهمیت آن را نمی پذیریم. اما بدون آن، ما نمی توانیم زمینه سازی کنیم.

به عنوان مثال، اگر کلمه بانک را در ذهن خود ببینیم، ممکن است به یک موسسه مالی یا مکانی که در آن اهدای خون ذخیره می شود، یا حتی یک باتری قابل حمل فکر کنیم. اما اگر جمله «برای درخواست وام به بانک می روم» را بخوانیم بلافاصله به بانکی که اشاره شده است پی می بریم. این به این دلیل است که ما به طور ضمنی به چند سرنخ توجه کردیم. از قسمت "رفتن به" فهمیدیم که بانک در این زمینه مکانی است و از قسمت "درخواست وام" دریافتیم که می توانید در آنجا وام دریافت کنید.

کل جمله اطلاعاتی را ارائه می دهد که با هم جمع می شوند تا تصویری ذهنی از چیستی بانک ایجاد کنند. فرض کنید یک ماشین می تواند همان کاری را که ما انجام می دهیم انجام دهد. در این صورت، بسیاری از مشکلات مهم پردازش زبان طبیعی مانند کلمات با معانی متعدد، جملات با ساختارهای دستوری متعدد و عدم اطمینان در مورد آنچه که یک ضمیر به آن اشاره دارد، حل خواهد شد.

## ۱-۲. دلیل استفاده از Multi-head attention به جای Single-head attention

Multi head attention به شبکه عصبی اجازه می دهد تا ترکیب اطلاعات را بین قسمت های مختلف یک دنباله ورودی کنترل کند، که منجر به ایجاد نمایش های غنی تر می شود، که به نوبه خود باعث افزایش عملکرد در وظایف یادگیری ماشین می شود.

برای غلبه بر برخی از مشکلات استفاده از self attention، از multi-head attention استفاده می شود. برای مثال جمله - "Bark is very cute and he is a dog". در اینجا، اگر کلمه "dog" را در نظر بگیریم، از نظر دستوری متوجه می شویم که کلمات "Bark"، "cute" و "he" باید با کلمه "dog" اهمیت یا ارتباط داشته باشند. این کلمات می گویند که نام سگ، bark است، سگ، نر است و سگ بامزه ای است. فقط یک مکانیسم توجه ممکن است نتواند به درستی این سه کلمه را به عنوان مرتبط با "سگ" تشخیص دهد، و می توانیم بگوییم که سه توجه در اینجا بهتر توجه سه کلمه را با کلمه "سگ" نشان می دهند. این امر بار روی یک توجه را برای یافتن همه کلمات مهم کاهش می دهد و همچنین شانس یافتن کلمات مرتبط تر را به راحتی افزایش می دهد.

بنابراین لایه های خطی بیشتری را به عنوان keys، queries و values کنیم. این لایه های خطی به صورت موازی آموزش می بینند و وزن های مستقلی نسبت به یکدیگر دارند. بنابراین اکنون، هر

یک از مقادیر، کلیدها و کوئری ها به جای یک خروجی، سه خروجی به ما می دهد. این ۳ کلید و پرس و جو اکنون سه وزن مختلف می دهند. این سه وزن سپس با ضرب ماتریس با سه مقدار، سه خروجی چندگانه به دست می دهند. این سه بلوک توجه در نهایت به هم متصل می شوند تا یک خروجی توجه نهایی را ارائه دهند.

## ۲-۱. پیاده سازی مدل BERT

مانند قسمت قبل کدهای بخش BERT، در فایل داده شده تکمیل شده است و در پیوست ارائه شده است.

### ۱-۲-۱. توضیح درمورد segment embedding در BERT

BERT قادر به حل مسئله NLP است که شامل طبقه بندی متن با توجه به یک جفت متن ورودی است. نمونه ای از چنین مشکلی طبقه بندی اینکه آیا دو قطعه متن از نظر معنایی مشابه هستند یا خیر. جفت متن ورودی به سادگی به هم متصل شده و به مدل وارد می شود. بنابراین BERT چگونه ورودی های یک جفت معین را Segment Embeddings با متمایز می کند. "Segment embeddings" اساساً شماره جمله ای است که در یک بردار کدگذاری می شود.

در BERT، مدل باید بداند که آیا یک نشانه خاص متعلق به جمله A است یا جمله B. این امر با تولید نشانه ثابت (fixed token) دیگری به نام «Segment embedding» - یک نشانه ثابت برای جمله A و یکی برای جمله B به دست می آید.

فقط دو نمایش برداری در لایه Segment Embeddings وجود دارد. همه نشانه های متعلق به ورودی ۱ به بردار اول (شاخص ۰) اختصاص داده می شوند، در حالی که همه نشانه های متعلق به ورودی ۲ به بردار دوم (شاخص ۱) اختصاص داده می شوند.

#### الگوریتم:

- اگر کلمه متعلق به جمله اول باشد ۰
- اگر کلمه متعلق به جمله دوم باشد ۱

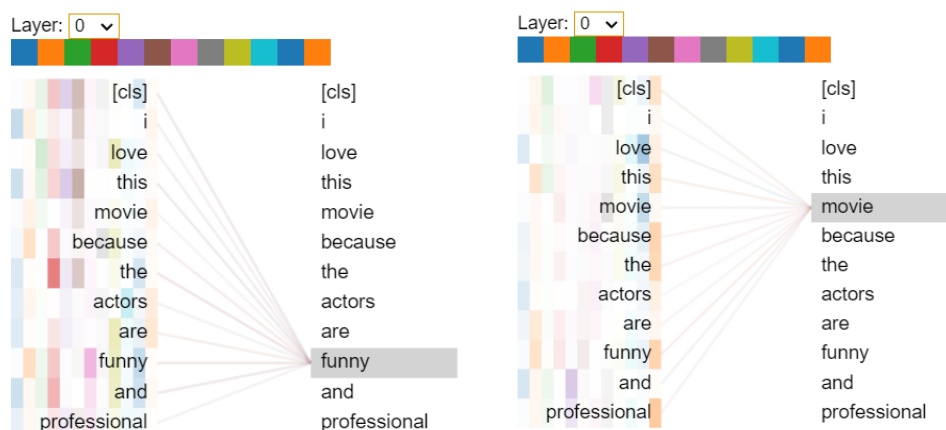
برای مثال به این دنباله نگاه کنید:

جدول ۱ - جدول نمونه segment embedding

[CLS]	I	LIKE	CATS	[SEP]	I	LIKE	DOGS
0	0	0	0	0	1	1	1

## ۲-۲-۱. نمونه خروجی در رابطه با یک فیلم

در این مدل جمله I Love this movie because the actors are funny and professional را به مدل دادیم و خروجی بصورت زیر میباشد که برای هر کلمه مشاهده میشود که ارتباط (توجه) آن با هر جمله مشخص میباشد.



شکل ۱ - نمونه خروجی شبکه BERT و مقدار توجه بر روی دو کلمه مختلف