|  |  |  |
| --- | --- | --- |
|  | **به نام خدا**  **دانشگاه تهران**  **دانشکده‌ مهندسی برق و کامپیوتر** |  |
| **درس شبکه‌های عصبی و یادگیری عمیق**  **تمرین پنجم** | | |

|  |  |
| --- | --- |
| نام و نام خانوادگی | **سارا رستمی – امین شاهچراغی** |
| شماره دانشجویی | **810۱00۳۵۵ - 810۱۹۹۱۹۶** |
| تاریخ ارسال گزارش | **۱۴۰۱.۱۰.۱۵** |

**فهرست**

[**پاسخ 1**. **آشنایی با مفهوم توجه و پیاده‌سازی مدل BERT** 4](#_Toc123757932)

[۱-۱. پیاده‌سازی کدگذار 4](#_Toc123757933)

[۱. 4](#_Toc123757934)

[۲. 4](#_Toc123757935)

[۲-۱. پیاده‌سازی مدل BERT 4](#_Toc123757936)

[۱. 4](#_Toc123757937)

[۲. 4](#_Toc123757938)

[**پاسخ ۲**. **آشنایی با مفهوم تبدیل‌کننده‌ها در تصویر** 5](#_Toc123757939)

[۱-۲. **آشنایی با مدل BEIT** 5](#_Toc123757940)

[۲-۲. **تقسیم‌بندی معنایی تصویر** 5](#_Toc123757941)

[۳-۲. **طبقه‌بندی تصاویر** 5](#_Toc123757942)

[۴-۲. **پرسش‌ها** 5](#_Toc123757943)

[۱. 5](#_Toc123757944)

[۲. 6](#_Toc123757945)

[۳. درستی یا نادرستی 6](#_Toc123757946)

**شکل‌ها**

شکل 1- خروجی مدل BERT برای جمله مذکور 5

شکل 2- خروجی مدل BERT برای جمله مذکور (ادامه) 5

شکل 3- خروجی مدل BERT برای جمله مذکور(ادامه) 5

**جدول‌ها**

No table of figures entries found.

# **پاسخ 1**. **آشنایی با مفهوم توجه و پیاده‌سازی مدل BERT**

۱-۱. پیاده‌سازی کدگذار

### ۱.

به فرآیندی که در آن بر روی قسمتی از ورودی وزن بیشتر و بر روی قسمتی وزن کمتر اعمال می شود، attention گفته می شود. لایه توجه در مدل BERT سه پارامتر Query Key, و Value را به عنوان ورودی می‌گیرد و یک تبدیل خطی را برای تولید دینامیک وزن‌ها برای connectionهای مختلف آغاز می‌کند. و سپس این تبدیل خطی را به یک ضرب داخلی scale شده می‌دهد.

در فرمول (۱)، K نشان‌دهنده‌ی Key، V نشان‌دهنده‌ی Value و Q نشان‌دهنده‌ی Query می‌باشد.

### ۲.

Multi-head attention این قابلیت را دارد که همزمان به چند پوزیشن متمرکز شود در حالی که می تواند sparsity را نیز حفظ کند. در واقع این نوع توجه این قابلیت را به مدل می‌دهد که دسته‌ی گسترده‌تری از روابط بین کلمات را capture کند(نسبت به single-head).

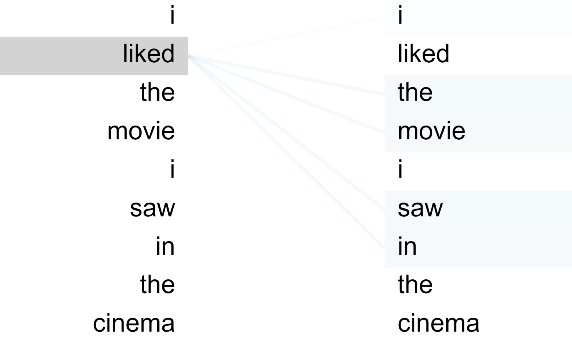
## ۲-۱. پیاده‌سازی مدل BERT

### ۱.

در segment embedding در واقع شماره جمله مورد نظر یا محل واقع شدن آن به مقدار عددی(بردار عددی) تبدیل می شود. به طور مثال در BERT مدل باید بداند که یک token به جمله‌ی A تعلق دارد یا جمله‌ی B. این هدف با تولید یک token ثابت (fixed) به نام segment embedding انجام می‌شود ( یک توکن ثابت برای جمله‌ی A و یک توکن ثابت برای جمله‌ی B). پو بردار بازنمایی در لایه‌ی segment embedding وجود دارد. توکن‌های متعلق به ورودی ۱ به اولین بردار assign می‌شوند (index 0) و توکن‌های متعلق به ورودی ۲ به دومین بردار assign می‌شوند (index 1).

### ۲.

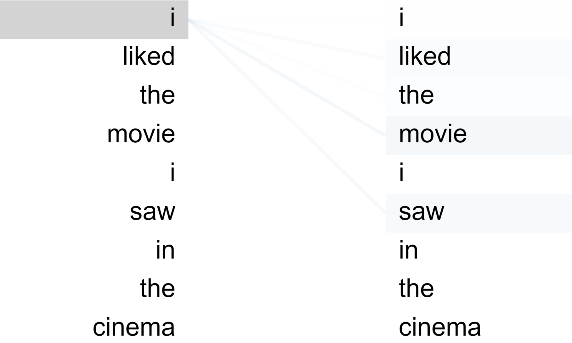
جمله “I liked the movie I saw in the cinema” به دسته بند داده شد و مقادیر لایه attention آن استخراج شد. وزن های attention در انتهای کد نمایش داده شده اند. به طور مثال بیشترین وزن کلمه “movie” مربوط به کلمات “I” و “cinema” است. خروجی مدل برای این جمله را در شکل‌های ۱ و ۲ و ۳ مشاهده می‌کنید.



شکل 1- خروجی مدل BERT برای جمله مذکور



شکل 2- خروجی مدل BERT برای جمله مذکور (ادامه)

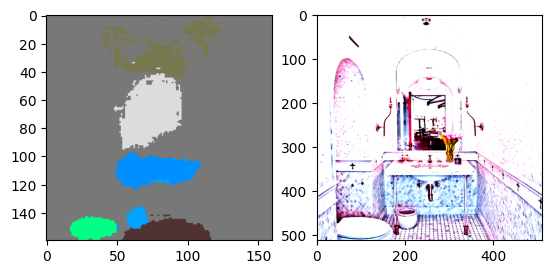
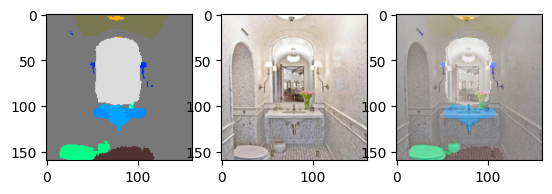


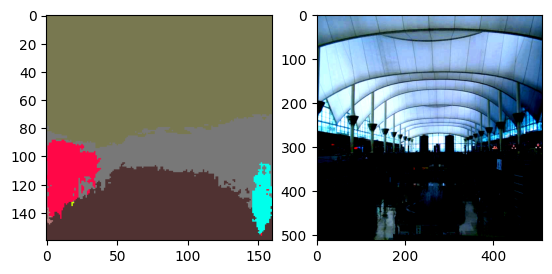
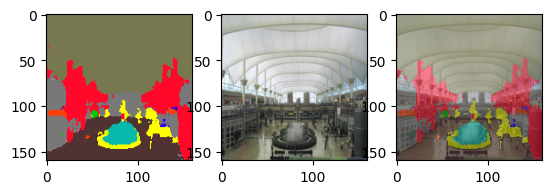
شکل 3- خروجی مدل BERT برای جمله مذکور(ادامه)

# **پاسخ ۲**. **آشنایی با مفهوم تبدیل‌کننده‌ها در تصویر**

## ۱-۲. **آشنایی با مدل BEIT**

## ۲-۲. **تقسیم‌بندی معنایی تصویر**





## ۳-۲. **طبقه‌بندی تصاویر**

مدل‌هایی

## ۴-۲. **پرسش‌ها**

### ۱.

در CNN مفهوم scan کردن عکس توسط kernel با مفهوم attention قرابت دارد.

### ۲.

مفهوم local attention به دسته ای از داده ها اشاره دارد که در نزدیکی یکدیگر قرار دارند در حالی که global attention به همجواری داده ها توجه نمی کند و attention را بر روی تمام داده ها اعمال می کند. CNN نیز معادل حالت hard-attention است که local-attention را به صورت سخت گیرانه تر با kernel کوچک تر انجام می دهد.

### ۳. درستی یا نادرستی

۱. نادرست

۲. درست

۳. درست

۴. نادرست