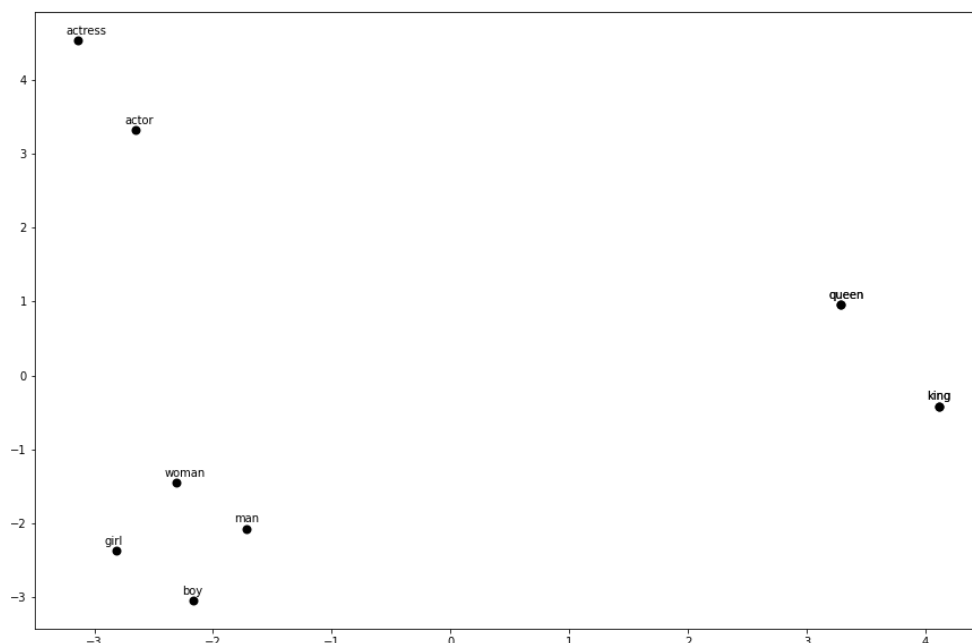


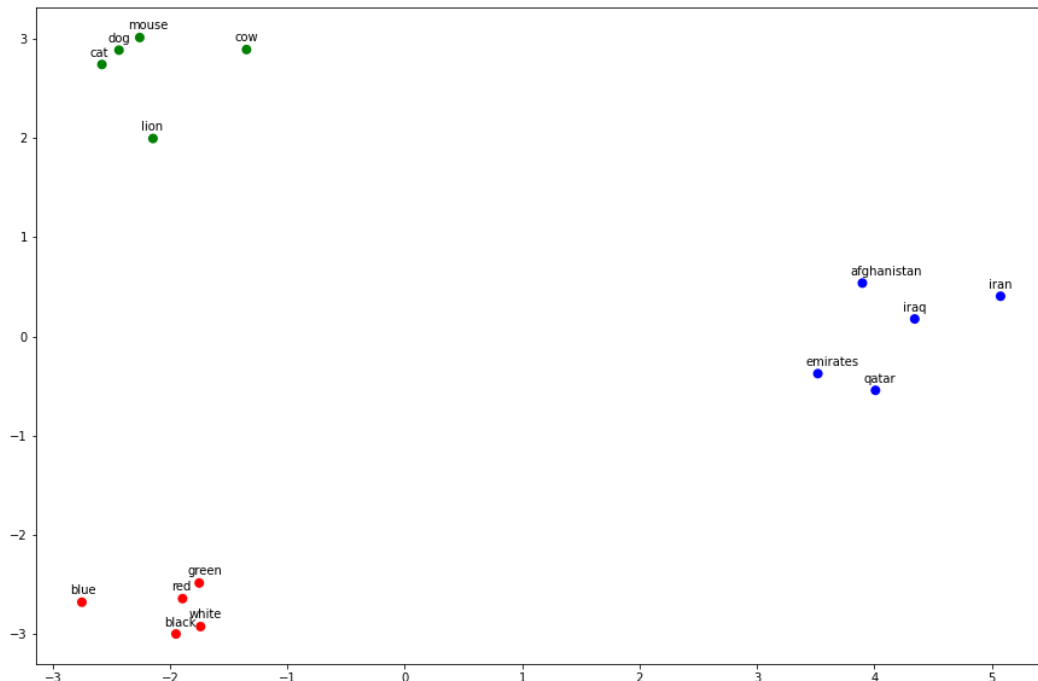


۱- در تمرین قبل با نحوه ساخت Word Embedding آشنا شدیم. اکنون می‌خواهیم با خواص بردارهای حاصل از Word Embedding بیشتر آشنا شویم. برای این کار در این تمرین از کتابخانه [gensim](#) استفاده می‌کنیم. برای حل این سوال لازم است مراحل زیر را به کمک نوت‌بوک موجود در پوشه تمرین انجام دهید.

- سوال ۱: در ابتدا تعدادی کلمه دلخواه را در قسمت مشخص شده در بخش Question 1 قرار دهید. خروجی این cell ۱۰ کلمه مشابه با هر کلمه را به همراه میزان شباهت دو کلمه نمایش می‌دهد. برداشت خود را از این خروجی بنویسید. این برداشت می‌تواند اشتباهات احتمالی خروجی یا هرگونه نکته قابل توجه در جدول خروجی باشد.
- سوال ۲: در این بخش لازم است تعدادی عضو شامل ۳ کلمه در قسمت تعیین شده در بخش Question 2 قرار دهید بطوریکه کلمه نخست، کلمه دوم و کلمه سوم در هر عضو به ترتیب کلمه‌ای دلخواه، کلمه‌ای که تصور می‌کنید نزدیک به کلمه انتخاب شده باشد و کلمه‌ای که تصور می‌کنید دور از کلمه انتخاب شده باشد می‌باشند. خروجی این cell شامل میزان شباهت کلمات نزدیک و دور انتخابی توسط شما به کلمه دلخواه انتخابی شما می‌باشد. تحلیل خود را از این میزان شباهت بیان کنید.
- سوال ۳: در این بخش (Question 3) تعدادی عضو شامل ۳ کلمه نوشته شده‌اند و هدف یافتن کلمه چهارم می‌باشد به گونه‌ای که تناسب بین کلمه سوم و کلمه دوم مشابه با تناسب کلمه اول و کلمه چهارم باشد. همانطور که مشاهده می‌کنید به ازای هر ۳ تایی از کلمات دو نوع فاصله محاسبه شده است، با ذکر دلیل توضیح دهید کدام فاصله برای یافتن کلمه چهارم مناسب است.
- سوال ۴: با اجرای این بخش (Question 4) توزیع کلماتی که از نظر جنسیت به دو گروه تقسیم می‌شوند را در قالب نمودار مشاهده می‌کنید (مشابه نمودار زیر). آیا می‌توان الگویی از این نمودار برداشت کرد؟ توضیح دهید.



- سوال ۵: مشابه بخش قبل، گروهی از کلمات را در قسمت مشخص شده در بخش Question 5 یادداشت نمایید بطوریکه این کلمات را بتوان به دو گروه تقسیم کرد. آیا الگویی مشابه با بخش قبل مشاهده می‌کنید؟ توضیح دهید.
- سوال ۶: در این بخش (Question 6) نموداری از توزیع ۳ گروه از کلمات شامل حیوانات، کشورها و رنگ‌ها را مشاهده می‌کنید (مشابه نمودار زیر). آیا می‌توان الگویی از این نمودار برداشت کرد؟ توضیح دهید.



- سوال ۷: مشابه بخش قبل، ۳ گروه از کلمات را در قسمت مشخص شده در بخش Question 7 یادداشت نمایید. آیا الگویی مشابه با بخش قبل مشاهده می‌کنید؟ توضیح دهید.
- سوال ۸: در این بخش (Question 8) ابتدا نشان داده شده است تناسب میان کلمات one و many مشابه تناسب میان کلمات cow و cows می‌باشد. اما مشکلی که وجود دارد عملکرد عجیب Word Embedding در یافتن کلمه‌ای است که بایستی در نتیجه یافتن کلمه مجهول در جمله "تناسب بین کلمه old با کلمه young مشابه با تناسب بین کلمه cow و کلمه مجهول می‌باشد" است. ابتدا بیان کنید کلمه مورد انتظار چه کلمه‌ای می‌باشد و سپس بوسیله دستوراتی که تا به اینجا آموختید نشان دهید چرا این کلمه به عنوان نتیجه نهایی انتخاب نشده است. این دستورات را در مکان مشخص شده در نوت‌بوک قرار دهید و دلیل خود را توضیح دهید.

۲- در این سوال قصد داریم یک شبکه Siamese برای یافتن embedding های تفکیک کننده برای اعداد موجود در مجموعه داده MNIST طراحی شود، شما در نوت‌بوک تهیه شده پیاده سازی های لازم را انجام خواهید داد.

وظایف شما:

الف. تابع triplet loss را با توجه به مطالب گفته شده در کلاس کامل کنید.

ب. یک شبکه برای ساخت embedding های ورودی طراحی کنید.

پ. در نهایت کل شبکه را آموزش دهید.

(برای درک بهتر مراحل، نوت بوک و اسلاید های مربوطه را مطالعه فرمایید)



خروجی نهایی خود را که embedding ها را با استفاده از t-SNE در فضای دو بعدی نمایش می دهد (این قسمت پیاده سازی شده) تحلیل کنید. (از t-SNE استفاده میشود تا ابعاد بردار ورودی را به ۲ کاهش دهیم تا در صفحه قابل نمایش باشد و این کار را هم برای embedding های استخراج شده و هم برای خود تصاویر خام انجام می دهیم).

همچنین توضیح دهید آیا در انتخاب سه تایی ها (tripple)، بهترین ها انتخاب شده اند؟ اگر نه توضیح دهید انتخاب بهینه سه تایی ها در این سوال چگونه است؟ (نیازی به پیاده سازی نیست).

۳- در این سوال شما با یک دیتاست کوچک پزشکی غیر متعادل (imbalanced) به نام derma mnist کار خواهید کرد. هدف دیتاست دسته بندی ده نوع مختلف از ضایعات پوستی می باشد و توضیحات کامل دیتاست و غیر متعادل بودن آن در فایل نوت بوک پیداست.

یک مدل کانولوشنی غیر خطی را با استفاده از keras functional api طراحی کنید و روی این دیتاست آموزش دهید.

الف) توضیح دهید آیا دقت معیار مناسبی برای ارزیابی عملکرد مدل می باشد؟

ب) معیارهای precision, recall, f1-score را محاسبه کنید، و توضیح دهید آیا مدل عملکرد خوبی داشته است یا نه؟

پ) نمودار precision, recall را رسم کنید و مفهوم AUC را توضیح دهید.

ت) Confusion Matrix را رسم کنید و مفهوم آن را با استفاده از مفاهیم TP, TN, FP, FN بیان کنید و دوباره با استفاده از این ماتریس بیان کنید آیا مدل عملکرد خوبی داشته است یا نه؟

ث) برای این مثال بهترین معیار ارزیابی شبکه به نظر شما چیست؟

ج) امتیازی: سعی کنید با استفاده از روش های موجود با مشکل غیر متعادل بودن دیتاست مقابله کنید و عملکرد مدل را بهبود دهید.

۴. هدف این سوال آشنایی با ابزار keras tuner می باشد.

در نوتبوک سوال، یک مدل ساده روی دیتاست fashion_mnist آموزش داده شده است. وظیفه شما بهینه کردن پارامتر های زیر در مدل موجود در نوت بوک با استفاده از keras tuner است:

- بهینه ترین تعداد نوروهای لایه dense در بازه 16 تا 512
- بهینه ترین نرخ یادگیری شبکه بین پارامترهای 0.0001, 0.0005, 0.001, 0.005, 0.01

در نهایت باید بهینه ترین پارامتر ها را گزارش کنید همچنین توضیح دهید چرا این پارامتر ها از پارامتر های قبلی بهتر شبکه را آموزش میدهند.



نکات تکمیلی:

1. لطفاً پاسخ سوالات (تئوری و توضیحات پیاده‌سازی) را به طور گویا و به زبان فارسی و در صورت امکان تایپ همراه با سورس کدهای نوشته شده، فقط در یک فایل فشرده شده به شکل HW14_YourStudentID.zip قرار داده و بارگذاری نمایید.
2. منابع استفاده شده را به طور دقیق ذکر کنید.
3. برای سهولت در پیاده‌سازی‌ها و منابع بیشتر، زبان پایتون پیشنهاد می‌شود. لطفاً کدهای مربوطه را در فرمت ipynb ارسال نمایید و هر کدام از موارد خواسته شده در بالا را در یک سلول جدید پیاده‌سازی نمایید.
4. ارزیابی تمرین‌ها براساس صحیح بودن راه حل‌ها، گزارش مناسب، بهینه بودن کدها و کپی نبودن می‌باشد.
5. در مجموع تمام تمرین‌ها، تنها 72 ساعت تاخیر در ارسال پاسخ‌ها مجاز است اما پس از آن به صورت خطی از نمره شما کسر خواهد شد (معادل با روزی 50 درصد).
6. تمرین‌ها باید به صورت انفرادی انجام شوند و حل گروهی تمرین مجاز نیست.
7. پرسش و پاسخ در رابطه با تمرین‌ها را می‌توانید در گروه مربوطه مطرح کنید.

موفق و سربلند باشید