

Q1)

در این سوال قرار است با Fine-tuning , pretrained آشنا بشیم و همچنین با Hugging Face آشنا بشیم.

در قسمت اول داده را دانلود کرده و در کولب باز می کنیم و ان را در پانداز ریخته داده تست و آموزش را جدا می کنیم.

تبدیل دیتاست پانداز به دیتاست مخصوص Hugging Face از توکنایزر Bert-bae-cased استفاده می کنیم.

توکنایز به دیتاستمون مپ می کنیم.

مدل پیش آموزش دیده شده bert لود می کنیم.

ارگیومننت های آموزش را ست می کنیم.

مدل های مختلف ارزیاب نگاه می کنیم و ان هایی را که می خواهیم انتخاب می کنیم.

ورودی های بخش آموزش را بهش می دهیم

پس از آموزش داده تست را هم برای پیش بینی به مدل می دهیم.

روی حالت های مختلف ارزیابی می کنیم تا ببینیم مدل چقدر خوب عمل کرده است.

Q2)

این قسمت به embding خواهیم پرداخت.

اول مواردی که نیاز داریم نصب می کنیم و بعد از ان ها کتابخونه های مورد نیاز را بارگذاری می کنیم.

مدل مورد نیاز که این جا word2vec-google-news-300 می باشد را اضافه می کنیم.

کلمات موجود را در دیکشنری اضافه می کنیم.

در این مرحله از ما خواسته شده Cosine similarity

استفاده کنیم که من اول از صفر پیاده سازی کردم که متاسفانه سرعت خوبی نداشت و خیلی طول می کشید برای محاسبات به خاطر همین کامنت کردم و از سعی کردم موارد محاسباتیش از کتابخونه های موجود کمک بگیرم.

قسمت complete_analogy

در این قسمت دنبال بهترین کلمه هستیم که اگر مثلاً نسبت مرد به زن را بخواهیم مقایسه کنیم پسر با چی می‌شه به دختر می‌رسیم که ما با استفاده از بردارهای کسینوس به دختر خواهیم رسید.

در این مرحله از صفر `neutralize` را پیاده سازی کردیم و در قسمت بعد هم `equalize` را پیاده سازی کردیم.