

Q1.1: What is the main advantage of word embeddings over one-hot encoding?

مزیت اصلی Word Embedding این است که می‌تواند شباهت و معنی کلمات را در قالب بردارهای عددی نشان دهد. در One-Hot، هر کلمه کاملاً مستقل از بقیه نمایش داده می‌شود و مدل نمی‌تواند بفهمد که مثلاً «dog» و «cat» از نظر معنایی به هم نزدیک هستند. اما در Word Embedding، کلمات مشابه بردارهای نزدیکی دارند و این باعث می‌شود مدل روابط معنایی را بهتر یاد بگیرد و کارایی بالاتری داشته باشد.

Q1.2: Explain the "distributional hypothesis" in one sentence.

کلماتی که در متن‌ها در موقعیت‌ها و با همسایه‌های مشابه ظاهر می‌شوند، عموماً معنی‌های مشابهی هم دارند.

Q1.3: What problem does LSTM solve that vanilla RNN cannot? Name the mechanism.

شبکه‌های RNN معمولی (Vanilla RNN) از مشکل محو شدن گرادیان رنج می‌برند که باعث می‌شود نتواند وابستگی‌های بلندمدت در دنباله‌ها را یاد بگیرند.

مدل‌های LSTM این مشکل را با معرفی یک حافظه‌ی داخلی (memory cell) و مکانیزم گیت‌ها (gates) شامل گیت ورودی، فراموشی و خروجی (input, forget, and output gates) حل می‌کنند؛ این دروازه‌ها جریان اطلاعات را کنترل کرده و اجازه می‌دهند گرادیان‌ها در بازه‌های زمانی طولانی‌تر بهتر منتقل شوند.

Q1.4: What is the purpose of the embedding layer in a neural network?

لایه‌ی Embedding در یک شبکه‌ی عصبی، لایه‌ای تخصصی است که داده‌های گستته و دسته‌ای (مثل کلمات، شناسه‌ها یا دسته‌ها) را به بردارهای پیوسته با ابعاد کمتر تبدیل می‌کند. این تبدیل به شبکه‌های عصبی کمک می‌کند تا ورودی‌های دسته‌ای را مؤثرتر پردازش کنند، چون آن‌ها را به جای نمایش‌های پراکنده و بسیار بزرگ One-Hot، به صورت بردارهای چگال (Dense) نمایش می‌دهد. یکی از کاربردهای رایج لایه‌های Embedding در پردازش زبان طبیعی (NLP) است. برای مثال، در یک مسئله‌ی دسته‌بندی متن، ابتدا کلمات به شاخص‌های عددی تبدیل می‌شوند. سپس لایه‌ی Embedding این شاخص‌ها را گرفته و بردارهای چگال متناظر با آن‌ها را تولید می‌کند و این بردارها به لایه‌های بعدی مثل LSTM‌ها یا ترانسفورمرها داده می‌شوند. این روش نسبت به One-Hot پیچیدگی محاسباتی را کاهش می‌دهد و به مدل کمک می‌کند بهتر تعمیم دهد، چون می‌تواند از شباهت بین کلمات استفاده کند. مثلاً اگر مدل یاد بگیرد که «happy» و «joyful» مقدار های مشابهی دارند، می‌تواند دانشی که از یکی یاد گرفته را برای دیگری هم به کار ببرد.

Q1.5: Why do we use bidirectional LSTMs for text classification?

از bidirectional LSTMs استفاده می‌کنیم چون می‌تواند هم‌زمان از اطلاعات قبل و بعد هر کلمه در جمله استفاده کند و در نتیجه معنی جمله را بهتر درک کند. یک LSTM معمولی فقط متن را از چپ به راست می‌خواند، پس هر کلمه فقط با توجه به کلمات قبلی تفسیر می‌شود. اما Bidirectional LSTM دو دارد یکی از چپ به راست، یکی از راست به چپ. خروجی این دو با هم ترکیب می‌شود، بنابراین نمایش هر کلمه شامل زمینه‌ی قبل و بعد آن است. که کمک می‌کند به درک بهتر معنی جمله و دقیق‌تر در طبقه‌بندی متن.

Part 4:

C) Find and explain 2 examples where word embeddings show bias (e.g., gender, racial).

مدل‌های word embedding از متون تولیدشده توسط انسان یاد می‌گیرند، در متون تاریخی، "پزشک" بیشتر با مردان و "پرستار" بیشتر با زنان مرتبط شده است این موضوع کلیشه‌های جنسیتی را تقویت می‌کند وقتی در الگوریتم‌های استخدام استفاده شود، ممکن است به زنان متقاضی شغل‌های "مردانه" آسیب برساند.

ابزارهای غربالگری رزومه ممکن است زنان متقاضی مهندسی را رد کنند. سیستم‌های پیشنهاد شغل ممکن است پرستاری را به زنان و مهندسی را به مردان پیشنهاد دهند.

بررسی این موضوع در کد:

```
6. man : doctor :: woman : ?  
nurse (score: 0.78)  
physician (score: 0.71)  
doctors (score: 0.68)
```

مثال دوم: سوگیری نژادی / ملیتی

ملیت‌ها و کلیشه‌ها:

"مکزیکی" با "غیرقانونی" یا "مهاجر" مرتبط می‌شود.

کشورهای "آفریقایی" با فقر / بیماری مرتبط می‌شوند.

کشورهای "اروپایی" با ثروت / فرهنگ مرتبط می‌شوند.

علت سوگیری نژادی این است که داده‌های آموزشی اغلب دارای گزارش‌گیری biased هستند و متون تاریخی حاوی دیدگاه‌های نژادپرستانه قدیمی هستند. این مورد می‌تواند پیامدهای را به دنبال داشته باشد از جمله الگوریتم‌های پلیس پیش‌بینی کننده مناطق اقلیت‌نشین را بیشتر هدف قرار می‌دهند. سیستم‌های تأیید وام ممکن است متقاضیان اقلیت را در مضيقه قرار دهند. تحلیل احساسات ممکن است گویش آفریقایی-آمریکایی را منفی تر تفسیر کند.