**Positional Embeddings**

**عملکرد:**

از آنجا که مدل ترنسفورمر ساختار دنباله‌ای ندارد (برخلاف RNN)، نیاز است موقعیت هر کلمه در توالی به مدل داده شود. این کار با **افزودن موقعیت هر کلمه به بردارهای آن** انجام می‌شود.

**مثال:**

در جمله‌ی "I like apples" اگر فقط embedding معمولی استفاده شود، مدل نمی‌داند کلمه‌ی "I" در ابتدای جمله است. Positional embedding کمک می‌کند مدل بداند "I" اول، "like" دوم و "apples" سوم است.

**اثر در نتایج:**

در مدل اجرا شده، اگر Positional Embedding حذف شود یا ثابت در نظر گرفته شود، مدل دقتش به‌شدت افت می‌کند چون ترتیب توکن‌ها را درک نمی‌کند.

**Attention و تأثیر تعداد Head (Multi-head Attention)**

**عملکرد Attention:**

مدل توجه می‌کند که کدام کلمات در جمله برای پیش‌بینی مهم‌تر هستند. هر کلمه با بقیه کلمات وزن‌دهی می‌شود.

**مثال:**

در جمله "The cat sat on the mat", برای پیش‌بینی کلمه "sat"، مدل به "cat" توجه بیشتری می‌دهد.

**تأثیر تعداد Headها:**

تعداد head بیشتر → مدل می‌تواند از دیدگاه‌های مختلف به جمله نگاه کند (یادگیری روابط متنوع‌تر).

* در اجرای من، وقتی num\_heads=1 انتخاب شد، دقت مدل پایین‌تر بود.
* وقتی num\_heads=4 یا 8 شد، مدل بهتر یاد گرفت چون می‌تواند چند نوع توجه موازی اجرا کند (مثلاً یک head روی زمان تمرکز کند، یکی روی جنس کلمه و...).

**تفاوت Attention و Self-Attention**

|  |  |
| --- | --- |
| **مفهــوم** | **توضیح** |
| Attention | توجه بین دو دنباله مختلف. مثلاً در ترجمه ماشینی، توجه از دنباله زبان مبدأ به زبان مقصد است. |
| Self-Attention | توجه بین عناصر یک دنباله. هر کلمه در جمله به خودش و بقیه کلمات نگاه می‌کند. |

**مثال:**

در جمله "She saw herself in the mirror"، در self-attention مدل می‌فهمد "herself" به "She" ارجاع دارد، چون به همه کلمات جمله نگاه می‌کند.

**در مدل کتاب:**

تمام attentionها از نوع **self-attention** بودند، چون مدل تنها یک دنباله را می‌بیند (نه دنباله مبدا و مقصد).

**نتیجه‌گیری**

* **Positional Embedding** برای دادن مفهوم ترتیب ضروری است.
* **Multi-head Attention** با افزایش تعداد headها می‌تواند رابطه‌های متنوع‌تری را یاد بگیرد.
* **Self-Attention** کمک می‌کند مدل مفهوم وابستگی بین توکن‌های همان جمله را بفهمد، بر خلاف Attention ساده که بین دو جمله است.

من تمام کد های کتاب مربوط به تمرین را پیاده سازی کردم ولی چون زمان اجرا هر epoc حدود 1 ساعت طول میکشد برای آن فقط یک ایپاک اجرا کردم.