

سوال اول

a.۱

گزینه ب به مقدار خروجیش نا مفهوم بود معنی 0.1 برای نشان دادن احساسات مثبت یا منفی اما به هر حال خروجی متن یا صوت یا ویدو نیست که طولش مشخص نباشد و خروجی یک چیز است.
گزینه ب و ج درست هستند. چون ورودی ورودی متن یا کلیپ صوتی یک دنباله با طول نامشخصه و many محسوب میشه و خروجی هم یک برچسبه و طول مشخصی داره و one محسوب میشه.

b.۱

ج) طبق صورت سوال حال و هوای گربه فقط به گذشته و حال تاثیر داره و آنچه بعدا اتفاق میفته رو نباید در نظر گرفت برای همین نباید از rnn دوطرفه استفاده کنیم. حجم محاسبات برای rnn دو طرفه زیاده و بیشتر طول میکشه و ارزششو نداره. د درست نیست چون میکه ۷ فقط به X بستگی داره و گذشته رو در نظر نمیگیره.

سوال دوم:

سوال سوم:

۳. لحاظ توجه softmax

مقادیر scores را با استفاده از ضرب داخلی ماتریس q در keys بدست آورده و
argmax روی آن اعمال می‌کنیم. خروجی ۱ به معنی اینست که values از آن است که
score مشاهده آن بیشترین باشد.

$$scores = keys^{-1} \cdot q$$

$$score_2 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & -2 & -4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 & 3 & 0 & 6 \end{bmatrix}$$

$$\alpha(k, q) = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$$

$$ff = \alpha \cdot values = \begin{bmatrix} 6 \\ 1 \\ 2 \end{bmatrix}$$

۳-ب

استفاده از عمل argmax در مدل‌هایی که از مکانیزم توجه (attention mechanism) استفاده می‌کنند، می‌تواند بر توانایی آموزش این مدل‌ها تأثیر بگذارد.

وقتی از argmax برای انتخاب کلیدها یا پرس‌وجوهای مورد استفاده در مکانیزم توجه استفاده می‌شود، این انتخاب به صورت گسسته و ناگهانی صورت می‌گیرد، به این معنا که فقط یکی از کلیدها یا پرس‌وجوها انتخاب می‌شود. انتخاب گسسته می‌تواند به مسائلی منجر شود که

به نوعی از اطلاعات مفیدی که ممکن است در کلیدها یا پرس‌وجوها وجود داشته باشد، عبور نمی‌یابد و این انتخاب گسسته ممکن است به مدل کمک نکند که از تمامی اطلاعات مفید برای حل مسئله بهره‌برداری کند.

در ضمن، هنگام آموزش مدل‌هایی که از مکانیزم توجه استفاده می‌کنند، گرادیان‌ها از لایه‌های آخر به لایه‌های اول شبکه منتقل می‌شوند. انتخاب گسسته باعث می‌شود که گرادیان‌ها برای کلیدها یا پرس‌وجوهای دیگر به مدل منتقل نشوند و به‌طور مستقیم گرادیان مشخص شده برای کلید یا پرس‌وجوی انتخاب شده به لایه‌های قبلی منتقل شود. این موضوع ممکن است باعث از دست رفتن اطلاعات مهم و بهبود امکانات مدل شود.

برای بهبود توانایی آموزش مدل‌هایی که از مکانیزم توجه استفاده می‌کنند، می‌توان از روش‌های دیگری برای انتخاب کلیدها و پرس‌وجوها استفاده کرد که این انتخاب به صورت پیوسته و نه گسسته باشد، به‌طوری که مدل بتواند از تمامی اطلاعات موجود در کلیدها و پرس‌وجوها بهره‌برداری کند. به عنوان مثال، از مکانیزم‌هایی مانند *soft attention* یا *learnable attention* استفاده کرد که انتخاب کلیدها و پرس‌وجوها را به صورت پیوسته و بر اساس وزن‌های احتمالی انجام دهند و به مدل اجازه دهند از تمامی اطلاعات موجود در ورودی‌ها استفاده کند.