

گزارش تمرین سوم

فرنام منصوری

94109595

سوال 1-

اول اشاره شود که من کد را کمی تغییر دادم که در هنگام train شدن batch های متفاوت را لرن کند جای اینکه کل دیتا را یکجا لرن کند در واقع به جای فول بچ من دیتا را بچ بچ کردم و بعد از اینکه با 1000 تا دنباله ی به طول 10 با بچ های به طول 10 دیتا را train کردم محتوای را از فایل q1.in خواندم و جوابش را روی q1.out ریختم.

در فایل normal_nn من برای آموزش شبکه ی عادی با لایه ی نهان به طول 4 دنباله های زیادی به طول 5 با مقدار رندوم را آموزش دادم (مقدار y را برایشان حساب کردم) برای اینکه کامل دیتا های 5 تایی را learn کند باید خطای آموزش به 0 میل کند و اینکه و همینطور چندین تست به طول 5 نیز به آن دادم ولی بیت توازن را نتوانست تشخیص بدهد.

البته همین کار را برای rnn نیز انجام دادم یعنی دنباله ای رندوم به طول 5 به تعداد زیاد به آن دادم ولی خطای آموزش به 0 میل نمیکرد و داده های تست را درست پیشبینی نمیکرد که فکر میکنم مشکل در بهینه سازی درست است البته وقتی دنباله های به طول 5 را نمیتواند یاد بگیرند به طول 6 هم یاد نمیگیرند ولی به میانگین خطای آموزش کتری به کمک rnn تونستم برسم برای دنباله های به طول 5 پس احتمالا برای دنباله های به طول 5 نیز همینطور است.

سوال 2-

1- من چندین بار شبکه را ران کردم و نتیجه ی تست برای ران های متفاوت کمی متفاوت بود یکبار شبکه توانست کل دنباله را جز x وسط پیشبینی کند ولی برای تمام ران های اخیر بین a و b ها یک تعدادی x پیشبینی میکند

2- اکثر نورن ها در زمان 14 یا 15 که محل x است تغییر شکل می دهند در مورد اینکه cell_state چی را نگه میدارد مثلا میتوان گفت که نورن 0 تا 14 مقدارش 0 است بقیه اش خطی است که میتواند نشانه ی فاصله تا x باشد و نورن 1 خطی است میتوان نشانه ی طول دنباله باشد و نورن 7 درست برعکس نورن 0 است و میتواند نشانه ی این باد که چقدر مانده تا x همینطور نورن 8 کاملا شبیه پله است و میتواند جای x را تایین کند ولی البته تمام این حرف ها شهودی است ولی کاملا نشانه ی این که در خود حافظه دارد سلول ما موجود است

3- input gate در 15 امی که آخرین a است زیرا نباید از حافظه تاثیر بگیرد و باید مقدار جدیدی بگیرد cell state و output gate هم در 15 امی زیرا مقدار cell state باید به طور کامل به hidden برود

مقدار cell state در نزدیکی x شروع میکند به تغییر کردن و برای تغییر خروجی آماده میشود

سوال 5:-

1- روش توجه سخت یک جایگاه را در بین جایگاه های ممکن برای یک چیز را در نظر میگیرد ولی در روش توجه نرم توزیع احتمال جایگاه های ممکن برای یک چیز را حساب می کنیم در روش نرم میتوان از back propagation استفاده کرد و آموزش آن ساده تر است زیرا در آموزش سخت اکثر گرادین ها 0 می شود ولی در توجه سخت معمولا از روش های reinforcement learning استفاده میشود

2- به ازای هر ورودی یک hidden state داریم اگر بعد از این که تمام ورودی ها را خواندیم بخواهیم خروجی را بدهیم در واقع داریم با اطلاعات hidden state آخر در واقع خروجی را می سازد ولی میتوان با یک ترکیب خطی از لایه های hidden هر ورودی میتوان به عنوان ورودی به هر سلول بعد از این که تمام ورودی ها تمام شد بدهیم حالا برای این که چه ترکیب خطی استفاده کنیم کافی است با کمک توجه نرم میتوانیم ضریب hidden state i ام را برای خروجی ز ام

احتمال اینکه z در جایگاه i ام باشد میگیریم یعنی مثلا برای ترجمه احتمال اینکه کلمه y خروجی z ام ترجمه y ورودی i ام باشد میگیریم

- 3- برای خروجی مورد نظر با کمک توجه می فهمیم که کجای تصویر بیشتر اهمیت دارد مثلا برای تسک image captioning می توان rnn را اینگونه طراحی کرد که محلی که به کلمه مربوط است را هم گزارش کند بعد درون شبکه کانولوشنال در یکی از لایه های آخر کانولوشن میبینیم که آن قسمت تصویر به کدام تکه از لایه y کانولوشنال مربوط است و آن تکه را به عنوان ورودی می توانیم برای کارمان استفاده کنیم مثلا در image captioning آن قسمت را به عنوان ورودی rnn بدهیم یا اگر توجه نرم است هر قسمت را ضرب در احتمالش کنیم و با هم جمع کنیم.
- 1- توجه محلی در قسمت attention یک محل مشخص میکنیم و در اطاف آن محل را به عنوان ورودی شبکه استفاده میکنیم در مواردی مثل translation یک متن طولانی کاربرد دارد چون که ترجمه یک متن به دوروبر آن فقط بستگی دارد توجه سراسری هم همان توجه نرم عادی است که ضریب استفاده از تمام داده ها را احتمال این که در آن مکان باشد می گذاریم