



## یادگیری عمیق

پاییز ۱۴۰۳

استاد: دکتر فاطمی زاده

گردآورندگان: نیلوفر عباسی، امیرعباس افضلی، آرمین قوجه زاده

مهلت ارسال: دوشنبه ۳ دی

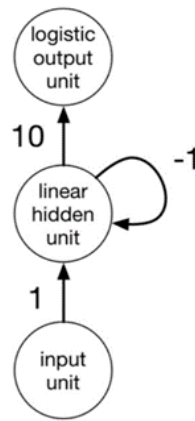
شبکه‌های بازگشتی و ترنسفورمر

تمرین چهارم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر پاسخ همه‌ی تمرین تا سقف ۵ روز و در مجموع ۱۵ روز، وجود دارد. پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند بود. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- همکاری و همفکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد. (دقت کنید در صورت تشخیص مشابهت غیرعادی برخورد جدی صورت خواهد گرفت.)
- در صورت همفکری و یا استفاده از هر منابع خارج درسی، نام همفکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.
- نتایج و پاسخ‌های خود را در یک فایل با فرمت zip به نام HW۴-Name-StudentNumber در سایت **CW** قرار دهید. برای بخش عملی تمرین نیز در صورتی که کد تمرین و نتایج خود را در گیت‌هاب بارگذاری می‌کنید، لینک مخزن مربوطه (repository) را در پاسخنامه خود قرار بدهید. دقت کنید هر سه فایل نوتبوک تکمیل شده بخش عملی را در گیت‌هاب قرار دهید. همچنین لازم است تا دسترسی‌های لازم را به دستیاران آموزشی مربوط به این تمرین بدهید.
- لطفا تمامی سوالات خود را از طریق صفحه درس در سایت **Quera** مطرح کنید (برای اینکه تمامی دانشجویان به پاسخ‌های مطرح شده به سوالات دسترسی داشته باشند و جلوی سوالات تکراری گرفته شود، به سوالات در بسترهای دیگر پاسخ داده نخواهد شد).
- دقت کنید کدهای شما باید قابلیت اجرای دوباره داشته باشند، در صورت دادن خطا هنگام اجرای کدتان، حتی اگر خطا بدلیل اشتباه تایپی باشد، نمره صفر به آن بخش تعلق خواهد گرفت.

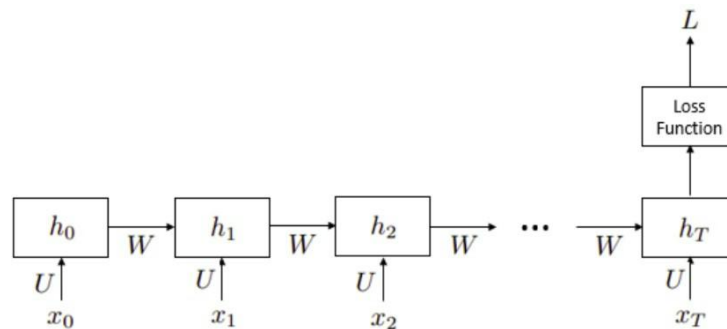
### سوالات نظری (۴۰۰ نمره)

۱. (۴۰ نمره) تعیین کنید شبکه زیر چه چیزی را تعیین می‌کند. (به طور دقیق‌تر، تابع محاسبه شده توسط واحد خروجی در مرحله زمانی نهایی را تعیین کنید.) سائیز خروجی‌ها مهم نیستند. همه بایاس‌ها صفر هستند. فرض کنید ورودی‌ها عدد صحیح هستند و طول دنباله ورودی زوج است. همچنین تابع فعال‌سازی را تابع سیگموئید در نظر بگیرید.



۲. (۴۰ نمره) دو مورد از مشکلاتی را که در استفاده از بردار one-hot برای بازنمایی کلمات وجود دارد را ذکر کنید.

۳. (۱۱۰ نمره) با توجه به شبکه عصبی بازگشتی شکل زیر به سوالات پاسخ دهید. دقت کنید که برای سادگی تمام مقادیر یعنی ورودی‌ها و وزن‌ها و خروجی مقادیر اسکالر هستند. همچنین فرض کنید تمام توابع فعالساز سیگموئید هستند.



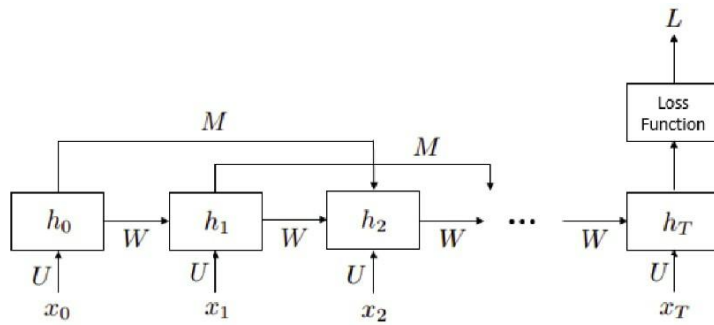
(آ) ابتدا گرادیان  $h_t$  یعنی  $\frac{\partial l}{\partial h_t}$  را بر حسب گرادیان  $h_{t+1}$  یعنی  $\frac{\partial l}{\partial h_{t+1}}$  بنویسید. ( $1 \leq t \leq T-1$ )

(ب) حال از رابطه قسمت قبل استفاده کرده و به شکل زنجیروار گرادیان  $h$  را بر حسب گرادیان  $h_T$  بنویسید.

حال می‌خواهیم روش‌هایی برای جلوگیری از محوشدگی و انفجار گرادیان را معرفی و تحلیل کنیم.

(آ) یکی از روش‌های مهم جلوگیری از محوشدگی و انفجار گرادیان مقداردهی اولیه صحیح وزن‌های شبکه است. توضیح دهید حداکثر مقدار اولیه  $W$  چند باشد تا فارغ از ورودی مطمئن باشیم که از همان ابتدا انفجار گرادیان رخ ندهد. (راهنمایی: یک حد بالا برای گرادیان  $h$  پیدا کنید.)

(ب) یکی از راه‌های جلوگیری از محوشدگی گرادیان استفاده از skip-connection است. شکل زیر را در نظر بگیرید که در آن هر  $h_t$  علاوه بر  $h_{t+1}$  به  $h_{t+2}$  هم متصل است. حال دوباره گرادیان  $h_t$  را بر حسب گرادیان  $h_{t+1}$  و  $h_{t+2}$  نوشته و توضیح دهید چرا این کار تا حد خوبی باعث کاهش اثر محوشدگی گرادیان می‌شود. ( $1 \leq t \leq T-2$ )

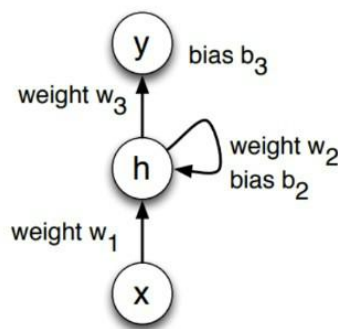


(ج) یکی از راه‌های جلوگیری از انفجار گرادین، برش گرادین است. که این خود به دو روش زیر راه‌حل برش توسط مقدار و برش توسط اندازه تقسیم می‌شود. این دو را جداگانه توضیح دهید. برتری برش توسط اندازه را به برش توسط مقدار را توضیح دهید.

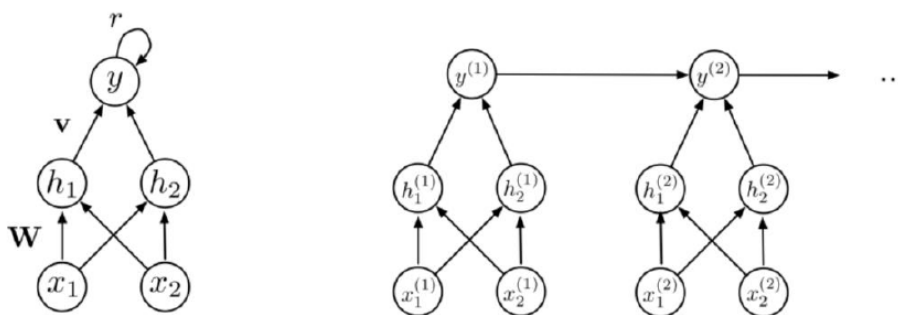
۴. (۷۰ نمره) در این مساله می‌خواهیم با مفاهیمی در تولید دنباله در شبکه‌های Seq2Seq و مزایا و معایب آن‌ها آشنا شویم. در این سوال می‌خواهیم مفهوم teacher forcing را بررسی کنیم. برای تولید دنباله ما می‌توانیم یک استراتژی خام اولیه در نظر بگیریم، می‌توان برای تولید توکن  $t + 1$  توسط رمزگشای زمان  $t + 1$ ، نشانه تولید شده توسط شبکه در زمان  $t$  را به عنوان ورودی به دیکودر در زمان  $t + 1$  بدهیم اما این حالت مشکلاتی دارد.

- (آ) ابتدا توضیح دهید این مشکلات چه چیزهایی هستند و سپس روش teacher forcing را توضیح داده و بگویید که teacher forcing چگونه این مشکلات را برطرف می‌کند.
- (ب) مشکل اصلی teacher forcing موضوعی به نام exposure bias است. این مشکل را توضیح دهید.
- (ج) یک از راه‌حل‌های مشکل exposure bias تکنیک scheduled sampling است، این تکنیک را توضیح داده و ذکر کنید این تکنیک چگونه باعث کاهش اثر exposure bias می‌شود.

۵. (۷۰ نمره) یک شبکه بازگشتی به صورت زیر را در نظر بگیرید. وزن‌ها و بایاس‌ها را به گونه‌ای تعیین کنید که در هر دنباله‌ای از اعداد تا زمانی که ورودی شبکه یک باشد، خروجی شبکه یک باقی بماند و به محض این که ورودی شبکه به صفر تغییر کند خروجی شبکه صفر شده و صفر باقی بماند. برای مثال خروجی شبکه به ازای ورودی ۱۱۱۰۱۰۱ برابر با ۱۱۱۰۰۰۰ می‌باشد.



۶. (۷۰ نمره) یک شبکه بازگشتی به صورت زیر در نظر بگیرید. فرض کنید این شبکه دو دنباله از اعداد صفر و یک را دریافت کرده و اگر دو دنباله برابر بودند عدد یک و در غیر این صورت عدد صفر را به عنوان خروجی برمی‌گرداند.



$$h^{(t)} = \phi(Wx^{(t)} + b)$$

$$y^{(t)} = \begin{cases} \phi(v^T h^{(t)} + r y^{(t-1)} + c) & \text{for } t > 1 \\ \phi(v^T h^{(t)} + c_0) & \text{for } t = 1, \end{cases} \quad \phi(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

ماتریس  $W$  یک ماتریس  $2 \times 2$  و  $b$  و  $v$  بردارهای دو بعدی و  $c$  و  $r$  و  $c_0$  مقادیر اسکالر می باشد. آن‌ها را به گونه‌ای تعیین کنید که شبکه کارکرد تعریف شده را داشته باشد. (راهنمایی: خروجی  $y^{(t)}$  در هر لحظه نشان می‌دهد آیا دو دنباله تا آن لحظه برابر بوده‌اند یا خیر. لایه مخفی اول نشان می‌دهد آیا دو ورودی در لحظه  $t$  صفر بوده‌اند یا خیر و لایه مخفی دوم نشان می‌دهد آیا دو ورودی در لحظه  $t$ ، یک بوده‌اند یا خیر)

## سوالات عملی (۶۰۰ نمره)

۱. (۱۵۰ نمره)

در این تمرین قصد داریم که یک شبکه عصبی بازگشتی ( $RNN$ ) را برای شناسایی احساسات توییت‌ها پیاده‌سازی کنیم. نوت‌بوک در نظر گرفته شده برای این تمرین شامل مراحل مختلفی است و لازم است با دقت آن‌ها را دنبال کنید و در بخش‌های مشخص شده کدهای مورد نیاز را تکمیل کنید. (لطفاً توجه فرمایید که در هر بخش لازم است توضیحات خود را در نوت‌بوک ارسالی به عنوان پاسخ وارد کنید و در بخش‌های مختلف ۴-۹ نوت بوک به توضیحات نمره اختصاص داده می‌شود. لازم است دستورالعمل‌های مرتبط با هر قسمت داخل نوت بوک را با دقت مطالعه کنید و سپس تغییرات لازم را ایجاد کنید.) هدف این تمرین یادگیری و درک عمیق‌تر *Recurrent Neural Networks* است.

در ادامه توضیح مختصری در مورد هر بخش از نوت بوک آورده شده است.

- (آ) **وارد کردن اطلاعات شخصی:** ابتدا باید اطلاعات شناسایی خود را در قسمت مشخص شده وارد کنید.
- (ب) **راه‌اندازی محیط و وارد کردن کتابخانه‌ها:** برای انسجام بیشتر و دسترسی آسان‌تر می‌توانید تمام کتابخانه‌های مورد نیاز خود را در این قسمت اضافه کنید و همچنین در طول تمرین از تابع‌های آماده‌ی نوشته شده در این قسمت برای نمایش نتایج خود استفاده کنید.
- (ج) **وارد کردن و تقسیم بندی داده‌ها:** در این بخش شما باید دیتاست خواسته شده را طبق دستورالعمل وارد نوت بوک کرده و آن را به مجموعه‌های آموزشی، اعتبارسنجی و آزمون تقسیم کنید.
- (د) **توکنیزاسیون توییت‌ها:** توییت‌ها را به توکن‌ها تبدیل کنید تا برای پردازش آماده شوند.
- (ه) **پدینگ و کوتاه کردن توالی‌ها:** طول توییت‌ها را بررسی کرده و توالی‌ها را به طول یکسان برسانید.
- (و) **تحلیل توزیع برچسب‌ها:** توزیع برچسب‌ها را بررسی کنید و در صورت وجود عدم تعادل، با استفاده از یک روش مناسب به آن رسیدگی کنید. در این قسمت لازم است روش مورد استفاده خود را داخل نوت بوک به طور کامل توضیح دهید.

(ز) **ایجاد مدل:** یک مدل  $RNN$  بسازید.

(ح) **آموزش مدل:** مدل خود را برای شناسایی احساسات آموزش دهید.

(ط) **ارزیابی مدل:** عملکرد مدل را بر روی مجموعه آزمون ارزیابی کرده و نتایج را تحلیل کنید.

۲. (۲۰۰ نمره) مدل‌سازی زبان یک تسک مهم در پردازش زبان طبیعی است که هدف آن تولید یک متن منسجم و معنادار بر اساس متن ورودی است. به تمپلیت سوال برای پیاده‌سازی و توضیحات تکمیلی مراجعه شود.

### ۱. دیتاست

(آ) از دیتاست WikiText2 استفاده کنید.

(ب) دیکشنری از روی داده آموزشی بسازید. (کم تکرارها را وارد دیکشنری نکنید.)

(ج) آماده‌سازی و پیش‌پردازش داده: تغییر شکل داده به شکل  $N \times B \times L$  یا  $M \times L$   
N تعداد کل بچه‌ها، B سائز بچه‌ها، L طول هر نمونه در بچه،  $M = B \times N$

(د) پیاده‌سازی ماژول فراخوانی داده: کاستوم دیتاست برای فراخوانی یک نمونه بچه/داده

### ۲. مدل

(ه) لایه امبدینگ +  $LSTM$  + لایه دسته‌بندی

### ۳. پیکره‌بندی

(و) تعریف مدل + تعریف تابع اتلاف + تعریف بهینه‌ساز + تنظیم هایپرپارامترها  
+ تعریف معیار ارزیابی (perplexity)

### ۴. آموزش و ارزیابی مدل

### ۵. به تولید هم نیاز داریم!

این کار در قالب تابعی با عنوان generate انجام می‌شود. حال تابعی بنویسید که:

(ز) متن ورودی کاربر را دریافت کند.

(ح) توکنایز و تبدیل به اندیس را انجام دهد.

(ط) به عنوان ورودی به مدل آموزش دیده وارد شود و خروجی طبقه بند دریافت شود. خروجی، برداری به طول تعداد توکن دیکشنری است.

(ی) خروجی از یک سافت مکس با Temperature مشخص رد شود. اگر  $\langle \text{unk} \rangle$  بود، یکبار دیگر انتخاب شود. و اگر به  $\langle \text{eos} \rangle$  رسیدیم، تابع generate پایان یابد.

(ک) توکن جدید به ورودی اضافه شود.

(ل) برگردیم به مرحله (ط)

۳. (۲۵۰ نمره) در این تمرین، هدف اصلی تبدیل یک مدل زبانی مولد مبتنی بر معماری Decoder-only به یک Encoder و ارزیابی آن در وظیفه طبقه‌بندی احساسات Sentiment Classification است. شما باید مدل را به گونه‌ای تغییر دهید که بتواند به عنوان یک Encoder عمل کرده و عملکرد آن را در پردازش‌های دوجته بهبود دهید. همچنین باید تاثیر تنظیمات مختلف در تولید متن را بررسی کنید و عملکرد مدل‌های مختلف در وظایف طبقه‌بندی را مقایسه و تحلیل کنید.

برای انجام این تمرین به نوتبوک EncoderLM.ipynb مراجعه کنید.