



تمرین سری پنجم
درس یادگیری عمیق

نام مدرس: دکتر محمدرضا محمدی

دستیاران آموزشی مرتبط: امیر جبلی، کسرا
شریعتی، آیدا خالقی، صادق پولایی

مهلت تحویل (بدون کسر نمره):
۸ دی ماه

سوالات تئوری (۵۰ نمره)

سوال ۱ - ۱۵ نمره

شبکه‌ای بازگشتی با دو لایه RNN و دو لایه FC به صورت زیر در نظر بگیرید (ورودی یک دنباله به طول ۵۰ با هر بردار ورودی ۶۴ بعدی):

Layer1: RNN(input_size=64, hidden_size=32, num_layers=1, nonlinearity='tanh',
batch_first=True)

Layer2: RNN(input_size=32, hidden_size=64, num_layers=1, nonlinearity='tanh',
batch_first=True)

Layer3: Linear(N = seq_len * 64, 512)

Layer4: Linear(512, 10)

الف) اندازه خروجی، تعداد پارامترها (وزن‌ها و بایاس‌ها جداگانه اعلام شود) هر لایه را با ذکر راه حل به تفکیک محاسبه کنید. (۶ نمره)

ب) تعداد کل عملیات ضرب و جمع (FLOPs) برای هر لایه را محاسبه کنید. (۵ نمره)

ج) اگر تغییرات زیر اعمال شود، اثر هر کدام را بنویسید: (۴ نمره)

۱. اگر طول دنباله دو برابر شود

۲. اگر سایز بردار ورودی به 128 تغییر کند

۳. اگر لایه دوم RNN به BiRNN تبدیل شود

۴. اگر لایه چهارم حذف شود

سوال ۲ - ۱۵ نمره

برای سوالات زیر پاسخ‌های مستدل ارائه دهید:

الف) برای هر نوع از داده‌های زیر مزایا و معایب هر کدام از شبکه‌های CNN و RNN را بنویسید. (۶ نمره)

نوع داده	RNN ¹		CNN	
	معایب	مزایا	معایب	مزایا
دنباله ساده ²				
تصویر				
صوت				
دنباله با روابط طولانی مدت ³				

ب) توضیح دهید چرا تابع فعال‌سازی تانژانت هایپربولیک (\tanh) در شبکه‌های بازگشتی یک انتخاب نسبتاً مناسب به‌شمار می‌آید. مزایا و محدودیت‌های آن را در مقایسه با سایر توابع فعال‌سازی بیان کنید. (۱ نمره)

ج) در معماری مدل‌های بازگشتی، روند محاسبات پیشرو (Forward Pass) را به‌صورت دقیق تشریح کنید. گراف محاسباتی⁴ زیر (تصویر ۱) را مشاهده کرده و محاسبات را مرحله‌به‌مرحله نشان دهید. (۳ نمره)

¹ منظور خانواده‌ی RNN هاست

² منظور دنباله‌ایست که اثر فراموشی در آن قابل چشم‌پوشی است

³ Long-term Dependency

⁴ Computational Graph

$$W_{xh} = \begin{bmatrix} 0.5 & -0.2 \\ -1.0 & 0.9 \end{bmatrix}$$

$$W_{hh} = \begin{bmatrix} 0.3 & 0.1 \\ -0.4 & 0.2 \end{bmatrix}$$

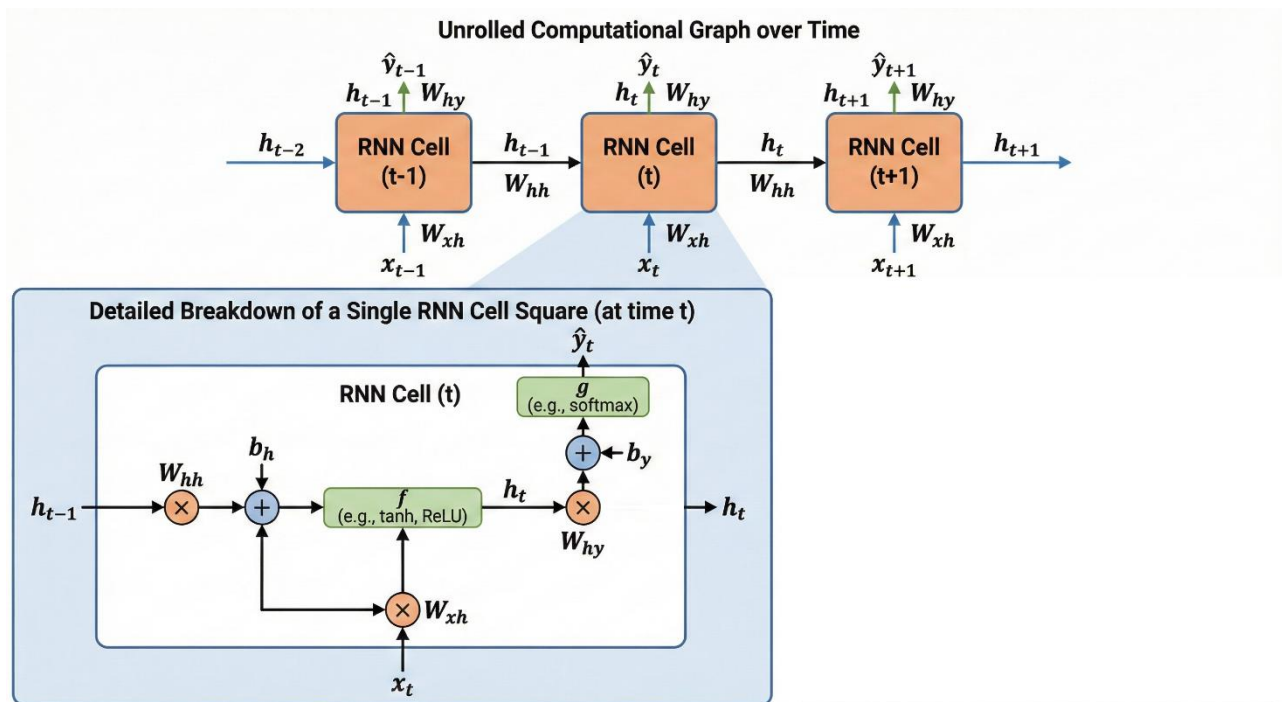
$$W_{hy} = \begin{bmatrix} 1.5 & -0.5 \\ 0.5 & 1.0 \end{bmatrix}$$

$$x_{t-1} = \begin{bmatrix} 0.8 \\ -0.5 \end{bmatrix}$$

$$h_{t-2} = \begin{bmatrix} -0.2 \\ 0.4 \end{bmatrix}$$

$$b_h = \begin{bmatrix} 0.1 \\ -0.1 \end{bmatrix}$$

$$b_y = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



تصویر ۱

د) توضیح دهید چرا شبکه‌های عصبی بازگشتی از نظر عملی امکان مدل‌سازی وابستگی‌های بلندمدت (Long-Term Dependencies) را فراهم نمی‌کنند؟ (۱ نمره)

ه) در ارتباط با معماری‌های مبتنی بر مدل‌های بازگشتی، در مورد حالات زیر توضیح دهید هر یک معمولاً در چه نوع مسائل و کاربردهایی مورد استفاده قرار می‌گیرند: (۱/۲۵ نمره)

- one-to-one
- one-to-many
- many-to-one
- many-to-many
- sequence-to-sequence

و) در مورد مفاهیم Vector Encoding و Vector Embedding شباهت‌ها و تفاوت‌های میان آن دو را از منظر نحوه‌ی نمایش داده، بعد فضا و یادگرفتنی بودن یا نبودن این نمایش‌ها، بیان کنید. (۱ نمره)

ن) کاربردها و مزایای استفاده از شبکه‌های بازگشتی دوسویه (Bidirectional RNN) را توضیح دهید. در چه نوع مسائل و سناریوهایی استفاده از BiRNN نسبت به RNN یک‌سویه برتری دارد؟ (۵/۰ نمره)

ح) برای جمع دودویی دو عدد n بیتی (bitwise addition)، تحلیل کنید که استفاده از یک MLP (شبکه‌ی پرسپترون چندلایه) مناسب‌تر است یا یک RNN با استدلال دقیق، مزایا و معایب هر معماری را در این مسئله بیان کرده و نتیجه‌گیری خود را توضیح دهید. (۱ نمره)

ط) توضیح دهید چرا در شبکه‌های بازگشتی کلاسیک، امکان بهره‌برداری کامل از ظرفیت موازی‌سازی پردازنده‌های گرافیکی (GPU) محدود است. (۲۵/۰ نمره)

سوال ۳ - ۱۰ نمره

به پرسش‌های زیر به‌صورت دقیق و مستند پاسخ دهید:

الف) سازوکار گیت‌ها^۵ در شبکه‌های حافظه طولانی-کوتاه‌مدت^۶ (LSTM) و واحدهای بازگشتی دروازه‌دار^۷ (GRU) را تشریح کنید. نقش هر یک از گیت‌ها را در فرآیند به‌روزرسانی حالت مخفی^۸ و کنترل جریان اطلاعات در توالی‌ها^۹ توضیح دهید. (۴ نمره)

ب) با تکیه بر تحلیل بخش (الف)، مبنای عملکردی و تفاوت‌های ساختاری LSTM و GRU را مقایسه کنید. (۳ نمره)

ج) تفاوت معماری و عملکرد ترکیب «CNN + LSTM» با معماری «ConvLSTM» را از نظر نحوه استخراج ویژگی، مدل‌سازی وابستگی‌های زمانی-مکانی^{۱۰} و سازمان‌دهی لایه‌ها توضیح دهید. (۳ نمره)

د) در شبکه‌های عصبی بازگشتی (RNN) مفهوم ابعاد کرنل (Kernel Size) و میدان تاثیر (Receptive Field) به چه صورت و با چه تعبیری قابلیت بازتفسیر دارند. (۳ نمره)

سوال ۴ (امتیازی) - ۱۰ نمره

سه سناریوی یادگیری عمیق زیر را با دیتاست‌های مشخص در نظر بگیرید :

^۵ Gate

^۶ Long Short-Term Memory

^۷ Gated Recurrent Unit

^۸ Hidden state

^۹ sequences

^{۱۰} spatiotemporal dependency modeling

– از بین معماری‌های Deep transition RNN, ConvLSTM, Skip RNN, Deep RNN برای هر سناریو، یک معماری را برگزینید و توضیح دهید چرا انتخاب شما است.

۱. تحلیل احساسات (Sentiment Analysis): روی جملات دیتاست IMDb که در آن وابستگی‌های بلندمدت¹¹ نقش کلیدی دارند.

مشکل چیست؟ در جملات یا متن‌ها، گاهی کلمه‌ای که در اول جمله آمده، معنی کلمه آخر جمله را تعیین می‌کند.

- مثال: من فکر می‌کردم فیلم خوبی باشد، اما در نهایت افتضاح بود.
- در اینجا کلمه "بود" به "من" وابسته است.
- در RNN‌های ساده، وقتی فاصله زمانی زیاد می‌شود، گرادیان در حین پس‌انتشار (Backpropagation) یا منفجر¹² می‌شود یا به سمت صفر میل می‌کند (Vanishing Gradient). در نتیجه مدل، ابتدای جمله را "فراموش" می‌کند.

۲. پیش‌بینی آبی بارش (Precipitation Nowcasting): روی توالی تصاویر راداری ابرها در دیتاست‌های هواشناسی، جایی که هم "مکان" ابرها و هم "زمان" حرکت آن‌ها حیاتی است.

مشکل چیست؟ فرض کنید یک توده ابر باران‌زا در شمال شهر قرار دارد و به آرامی به سمت شرق حرکت می‌کند. یک LSTM معمولی برای پردازش این ورودی، مجبور است تصویر دوبعدی را بکوبد و به یک بردار یک‌بعدی (Flatten) تبدیل کند. با این کار، رابطه همسایگی پیکسل‌ها نابود می‌شود! مدل دیگر نمی‌داند شمال و شرق کجای بردار است. این کار باعث می‌شود مدل اطلاعات مکانی (Spatial) را کاملاً از دست بدهد و نتواند مسیر حرکت ابر را روی نقشه درست پیش‌بینی کند.

۳. تشخیص کنش (Action Detection): در ویدیوهای طولانی و برش‌نخورده دیتاست ActivityNet، جایی که کنش‌ها بسیار پراکنده¹³ هستند و برای کاهش بار محاسباتی، مدل باید یاد بگیرد فریم‌های نامرتبط (Background) را پردازش نکند.

مشکل چیست؟ فرض کنید یک ویدیوی ۱۰ دقیقه‌ای از یک اتاق دارید که فقط در ثانیه ۵:۳۰ یک نفر وارد می‌شود و یک لیوان برمی‌دارد.

- یک RNN یا LSTM معمولی مجبور است تک‌تک فریم‌های خالی قبل و بعد از آن صحنه را پردازش کند. این یعنی میلیون‌ها ضرب ماتریسی بیهوده برای پردازش دیوار خالی!
- این کار هم زمان‌بر است و هم هزینه محاسباتی را بالا می‌برد.

¹¹ Long-term Dependency

¹² Exploding gradient

¹³ Sparse

سوال ۶ (۲۰ امتیاز)

قسمت نکات تکمیلی در انتهای فایل خوانده شود

در این تمرین عملی، شما به صورت تجربی درک خواهید کرد که چرا RNN های ساده در توالی های با وابستگی بلندمدت شکست می خورند. به جای اینکه صرفاً به شما گفته شود گرادیان ها محو می شوند، شما این پدیده را مستقیماً مشاهده خواهید کرد. برای این تمرین از این [دیتاست](#) استفاده کنید. در این تمرین سه آزمایش اصلی را انجام می دهیم.

آزمایش A: تأثیر طول توالی

هر دو مدل Simple RNN و LSTM را روی 4 طول مختلف آموزش دهید: 50, 100, 200, 400 توکن برای هر طول، مدل جداگانه ای آموزش دهید و دنباله ها را به طول مشخص truncate یا pad کنید. خروجی مورد نظر یک نمودار است که دقت هر یک از مدل ها را با طول مشخص نشان می دهد. این نمودار را با دقت بررسی و تحلیل کنید و یافته های خود را در گزارش ذکر کنید.

آزمایش B: اندازه گیری جریان گرادیان

از PyTorch hooks برای گرفتن گرادیان ها در طول backpropagation روی یک batch از validation set با طول 200 توکن استفاده کنید. برای هر گام زمانی، نرم L2 گرادیان را محاسبه کنید.

خروجی باید یک نمودار با مشخصات زیر باشد (برای Simple RNN و LSTM در کنار هم):

- محور افقی = موقعیت گام زمانی (1 تا 200)
- محور عمودی = بزرگی گرادیان (مقیاس لگاریتمی)

اندازه گرادیان ها را در timestep های مختلف مشاهده و مقایسه کنید و دریافت های خود را توضیح دهید.

آزمایش C: تحلیل پویایی hidden state

10 دنباله را به عنوان تست (5 مثبت و 5 منفی) همه با طول حداقل 150 توکن انتخاب کنید و بردارهای hidden state را در هر گام زمانی استخراج کنید. نرم L2 هر بردار hidden state را محاسبه کنید. خروجی نمودارهایی است که نشان می دهند نرم hidden state ها چگونه در طول گام های زمانی تکامل می یابد (به صورت کنار هم با رنگ های متفاوت برای دنباله های مثبت و منفی). برای LSTM، نرم cell state را نیز رسم کنید. این نمودار های را با دقت ارزیابی کرده و تمامی دریافت های خود را در گزارش خود بیاورید.

سوال ۷ (۴۰ امتیاز)

متن این سوال را در پوشه Practical 2 می توانید مشاهده کنید.

نکات تکمیلی:

برای سوال عملی اول یک نوتبوک تهیه شده است و می‌توانید در پوشه Practical مشاهده کنید. استفاده از نوتبوک یا پیاده سازی از پایه به عهده دانشجو است. تمام کدها بر اساس روند حل سوال و توسط خود دانشجو نوشته شود (استفاده از چت‌بات‌ها مجاز است، به شرط مرجع دادن و تسلط به پاسخ در ارائه).

دانشجویان محترم حتماً فایل قوانین را مطالعه کرده و در انجام و ارسال تمرین رعایت بفرمایید.

موفق و سربلند باشید.