

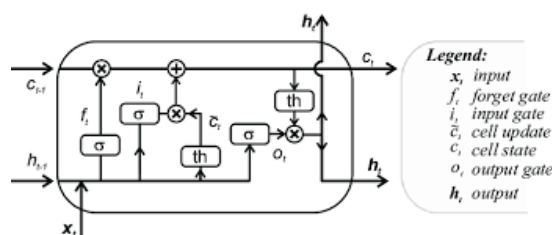
## شبکه‌های عصبی

### تمرین ششم

### سوال یک

همان‌طور که می‌دانیم هر دو سلول LSTM و GRU برای ساختن شبکه‌های بازگشتی استفاده می‌شوند. هر دو این شبکه‌ها نیز برای بهبود عملکرد واحد RNN عرضه شده‌اند. در ادامه جزئیات بیشتری از این دو مدل تشریح می‌شود.

واحد LSTM از سه دروازه<sup>۱</sup> اصلی تشکیل شده است: دروازه ورودی، دروازه صرف‌نظر<sup>۲</sup> و دروازه خروجی. هر یک از این دروازه‌ها وظیفه خاصی به عهده دارد. دروازه ورودی وظیفه تصمیم‌گیری درباره حجم اطلاعات وارد شده به واحد LSTM دارد. دروازه صرف‌نظر همان‌طور که از نام آن پیداست وظیفه تصمیم‌گیری درباره میزان دانش که باید دور ریخته شود را به عهده دارد. دروازه خروجی نیز خروجی واحد LSTM را بر اساس دانش کسب شده تولید می‌کند. در شکل ۱ نمایی از واحد LSTM آورده شده است.



شکل ۱: ساختار واحد LSTM

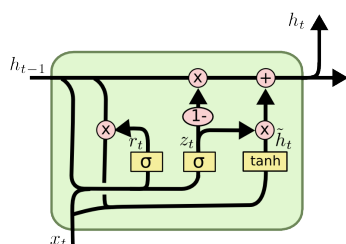
واحد GRU از دو دروازه تشکیل شده است: دروازه به‌روزرسانی<sup>۳</sup> و دروازه بازنشانی<sup>۴</sup>. عملکرد دروازه بازنشانی مشابه دروازه صرف‌نظر در واحد LSTM است. این واحد وظیفه تصمیم‌گیری در رابطه با فراموش کردن دانشی که مدل تا به حال یادگرفته را به عهده دارد. واحد به‌روزرسانی نیز مسئول تصمیم‌گیری در خصوص میزان دانشی است که مدل به واحد بعدی منتقل می‌کند. همان‌طور که مشاهده می‌شود عملکرد این دروازه بسیار شبیه عملکرد دروازه خروجی در واحد LSTM است. در شکل ۲ نمونه‌ای از واحد GRU آورده شده است. با توجه به توضیحات داده شده مشخص است که واحد GRU تعداد پارامترهای کمتری نسبت به واحد LSTM دارد. این مسئله باعث می‌شود قدرت شبکه GRU در مقایسه با LSTM کمتر شود. از طرفی بیشتر شدن قدرت باعث افزایش واریانس شده و در نتیجه احتمال بیش‌برازش واحد LSTM رو داده‌های ورودی بیشتر می‌شود. بنابراین اگر تعداد

<sup>۱</sup>gate

<sup>۲</sup>forget

<sup>۳</sup>update

<sup>۴</sup>reset



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

شکل ۲: ساختار واحد GRU

داده‌ها و ویژگی‌ها زیاد باشد بهتر است از مدل LSTM استفاده شود، در غیر این صورت استفاده از شبکه عصبی GRU منطقی‌تر به نظر می‌رسد. با همهی این تفاسیر توصیه‌ای که در استفاده از مدل‌های LSTM و GRU در حل مسائل می‌شود این است که هر دو مدل برای حل مسئله امتحان شود.

## سوال دو

در ابتدا توضیحاتی در رابطه با پیش‌پردازش انجام شده بر روی داده‌ها توضیحاتی داده می‌شود. در قدم اول فایل‌های csv که در اختیار ما قرار داده شده بود با استفاده از کتابخانه pandas بارگیری می‌شود. در این میان بعضی از فایل‌ها داده خاصی نداشته و خالی هستند بنابراین این فایل‌ها دور ریخته می‌شوند. سپس سطرهای مشترک موجود در جداول با استفاده از فیلد <DTYYYYMMDD> ادغام و مرتب می‌شود. پس از ادغام فایل‌ها داده‌های یکسال اخیر را در نظر گرفته و باقی داده‌ها را حذف می‌کنیم. همچنین ستون‌هایی که تعداد مقادیر غیرتکرار آن‌ها کم‌تر از ۲ است را حذف می‌کنیم. استدلال ما برای حذف این ستون‌ها این است که ستون‌هایی که تعداد مقادیر غیرتکراری آن‌ها بسیار کم است، بیشتر از آن که به مدل پیش‌بینی بهتر کمک کنند، باعث گمراهی آن می‌شوند. با اعمال این شرط ستون‌های <TICKER>، <COL12> و <COL13> از مجموعه داده‌ها حذف می‌شوند.

در نهایت لازم است که تعیین کنیم آیا شاخص در یک روز مشخص صعودی بوده یا خیر. برای این کار از تفاضل close هر روز نسبت به روز قبل استفاده می‌کنیم. اگر مقدار این تفاضل مثبت بود یعنی شاخص مثبت بوده و با مقدار یک نشان می‌دهیم در غیر این صورت مقدار شاخص منفی بوده و با عدد صفر نمایش می‌دهیم.

علاوه بر اقدامات بالا، کارهای دیگری نیز برای پیش‌پردازش داده‌ها صورت گرفته است:

- دو سطر از داده‌ها در بعضی از ستون‌ها بدون مقدار بودند. با توجه به آن که تعداد کل داده‌هایی که در اختیار داریم کم است، این دو سطر را دور نریختیم و میانگین سطرهای قبل و بعد از سطر مذکور را به عنوان مقدار آن خانه خالی در نظر گرفتیم.
- برای آن که به نتیجه بهتری برسیم با استفاده از MinMaxScaler ارائه شده توسط کتابخانه sklearn داده‌های هر ستون را با توجه به کمینه و بیشینه آن ستون، به بازه (0, 1) نگاشت می‌کنیم.
- برای آن که داده‌ها ماهیت سری زمانی داشته باشند، داده‌های هر ۱۰ روز متوالی را با یک برچسب به مدل دادیم. منظور از برچسب صعودی یا نزولی بودن شاخص در روز ۱۱ام پس از این ۱۰ روز است.

پس از پیش‌پردازش داده‌ها نوبت به تقسیم داده‌ها به مجموعه داده‌های آموزش، ارزیابی و آزمون می‌رسد. ما در این مرحله تلاش می‌کنیم داده‌ها را به نحوی تقسیم کنیم که تا جای امکان هر سه مجموعه داده متعادل باشند. پس از تقسیم داده‌ها در مجموعه داده آموزشی، ارزیابی و آزمون به ترتیب حدود ۵۷، ۵۶ و ۵۵ درصد داده‌ها را داده‌های با برچسب یک تشکیل می‌دهند. این درصدها بیانگر آن هستند که توزیع برچسب‌ها در هر سه مجموعه داده یکسان و متعادل هستند.

پیش از ارائه نتایج نیز لازم است نکاتی در رابطه با چگونگی استخراج نتایج بیان شود.

- نتایج با ۵ مرتبه آموزش و ارزیابی ارائه شده است. بنابراین می‌توان انتظار داشت که با تکرار آزمایش‌ها همین نتایج دوباره به دست آید.

- برای به دست آوردن نتایج مقادیر مختلف برای پارامترهای زیر بررسی شده و نتایج بر اساس بهترین پارامترها گزارش شده است.

– نرخ یادگیری. برای نرخ یادگیری مقادیر 0.01، 0.001 و 0.001 امتحان شده است.

– تابع بهینه‌ساز. توابع بهینه‌ساز Adam و SGD برای رسیدن به جواب بهینه بررسی شده است.

- تعداد گام یادگیری مدل بر روی ۳۰۰ تنظیم شده است.

- با توجه به زیاد بودن تعداد گام یادگیری برای جلوگیری از بیش‌برازش، از یک Early Stopping Callback با مقدار صبر ۲۰ بر روی خطای داده‌های ارزیابی استفاده شده است.

نتایجی که مدل LSTM می‌رسد در جدول ۱ آمده است. این نتایج با استفاده از بهینه‌ساز Adam و با نرخ یادگیری 0.001 به دست آمده است.

Table 1: نتایج عملکرد مدل LSTM با تعداد مختلف ابعاد لایه مخفی

# of hcells	Train Accuracy	Validation Accuracy	Test Accuracy
8	0.6	0.57	0.56
32	0.61	0.59	0.56
64	0.62	0.57	0.58
256	0.62	0.57	0.57
512	0.59	0.57	0.55

نتایج عملکرد مدل GRU با استفاده از بهینه‌ساز Adam و نرخ یادگیری 0.001 نیز در جدول ۲ مشاهده می‌شود.

Table 2: نتایج عملکرد مدل GRU با تعداد مختلف ابعاد لایه مخفی

# of hcells	Train Accuracy	Validation Accuracy	Test Accuracy
8	0.62	0.6	0.53
32	0.63	0.58	0.56
64	0.61	0.58	0.57
256	0.65	0.57	0.58
512	0.65	0.56	0.56

همان‌طور که مشاهده می‌شود، عملکرد هر دو شبکه LSTM و GRU با افزایش تعداد نورون‌ها در لایه‌های مخفی اندکی بهتر می‌شود. اما با ادامه افزایش تعداد نورون‌ها عملکرد شبکه افت می‌کند. به عبارت دیگر با افزایش تعداد نورون‌ها شبکه بهتر توانسته است الگوی کلی داده‌ها را شناسایی کرده و یاد بگیرد. اما این روند با افزایش بیش از حد نورون‌های لایه مخفی عکس شده و شبکه الگوهای بسیار جزئی از داده‌های آموزشی را یاد گرفته که در نتیجه نتوانسته روی داده‌های ارزیابی و آزمون به خوبی عمل کند. به عبارت بهتر با افزایش تعداد نورون‌ها پدیده بیش‌برازش رخ داده است. البته همان‌طور که پیش‌تر ذکر شد ما از یک Early Stopping Callback برای جلوگیری از بیش‌برازش مدل استفاده کرده‌ایم. اگر از این callback استفاده نمی‌شد تاثیر بیش‌برازش مدل بسیار واضح‌تر می‌بود.

با توجه به جدول، عملکرد هر دو شبکه LSTM و GRU تا حد زیادی شبیه یکدیگر هستند. معیاری که برای انتخاب بهترین عملکرد در نظر می‌گیریم عملکرد هر دو مدل روی داده‌های آزمون است. هر دو این شبکه‌ها توانسته‌اند به دقت ۵۸ درصد روی داده‌های آزمون دست یابند. اما هر یک از این دو شبکه با تعداد متفاوت نورون‌های لایه مخفی. شبکه LSTM با توجه به پیچیدگی زیادتری که دارد به تعداد نورون‌های کمتری نیاز داشته است تا به این عملکرد برسد در حالی که شبکه GRU برای آن که به این دقت برسد، به تعداد بسیار بیشتری نورون نیاز داشته است. نکته مشابهی در هنگامی که تعداد نورون‌های لایه مخفی کم است مشاهده می‌شود. هنگامی که تعداد نورون‌های لایه مخفی هر دو شبکه ۸ است، شبکه LSTM عملکرد بهتری را ارائه می‌دهد.

یکسان بودن بهترین عملکرد هر دو مدل را می‌توان کم بودن داده آموزشی در عین پیچیدگی آن در نظر گرفت. به نظر می‌رسد این عملکرد یک کران بالا برای هر دو مدل باشد. چرا که هر دو مدل با تنظیم درست تعداد پارامترها می‌توانند به آن برسند. برای آن که به نتایج بهتری از عملکرد مدل‌ها در این حالت دست پیدا کنیم حالت‌های بسیار زیادی تست شده است. نتایج ارائه شده با استفاده از بهترین پارامترها ارائه شده است. تلاش‌هایی که برای رسیدن به بهترین نتایج انجام شده است در ادامه آورده می‌شود. متأسفانه بیشتر این تلاش‌ها باعث بهبود مدل نشده‌اند.

- استفاده از بهینه‌سازهای مختلف
- استفاده از نرخ‌های مختلف یادگیری
- تغییر بازه دخیل در پیش‌بینی مدل
- تفکیک چندباره مدل به مجموعه داده‌های آموزش، ارزیابی و آزمون و آموزش مدل بر روی آن‌ها
- استفاده از شاخص کل به صورت مجزا برای آموزش مدل و حذف سایر شاخص‌ها

## سوال سه

نتایج عملکرد مدل پشته‌ای حاصل از سلول‌های LSTM در جدول ۳ و نتایج مدل پشته‌ای از واحدهای GRU در شکل ۴ دیده می‌شود. برای آموزش هر دو این شبکه‌ها از بهینه‌ساز Adam و نرخ یادگیری 0.0001 استفاده شده است.

همان‌طور که مشاهده می‌شود نتایج عملکرد هر دو مدل نسبت به حالتی که تنها از یک لایه سلول بازگشتی استفاده می‌شود اندکی بهتر است. با افزایش تعداد لایه‌های بازگشتی نیز عملکرد مدل‌ها اندکی بهبود می‌یابد. تفسیر این اتفاق آن است این مدل‌ها بهتر می‌توانند الگوهای پیچیده موجود در داده‌ها را شناسایی کنند. در نتیجه عملکرد بهتری روی داده‌های ارزیابی و آزمون دارند.

انتظار می‌رفت پشته کردن لایه‌های بازگشتی GRU عملکرد بهتری نسبت به LSTM داشته باشد. چرا که شبکه‌های GRU از ساختار ساده‌تری برخوردار هستند بنابراین پشته کردن که به معنای افزایش پارامترها است، قدرت آن‌ها را به طور معقول‌تری در مقایسه با پشته کردن شبکه‌های LSTM افزایش می‌دهد. به هر حال با پشته کردن هر دو مدل عملکرد هر دو مدل بهبود یافته است. شاید یکی از دلایلی که باعث جلوگیری از بیش‌برازش مدل‌های پشته‌ای LSTM شده است، استفاده از نرخ یادگیری بسیار کم (0.001) در کنار Early Stopping Callback باشد. با توجه به آن که تعداد گام آموزشی بسیار زیاد (۳۰۰ گام) است مدل GRU به آرامی به سمت بهتر شدن پیش می‌رود در حالی که آموزش مدل LSTM در همان گام‌های ابتدایی متوقف می‌شود.

Table 3: نتایج عملکرد مدل LSTM با تعداد مختلف ابعاد لایه مخفی و تعداد پشته‌های مختلف

# of stacks	# of hcells	Train Acc	Validation Acc	Test Acc
2	8	0.63	0.6	0.6
	32	0.65	0.6	0.6
	64	0.64	0.57	0.6
	256	0.61	0.61	0.59
	512	0.65	0.59	0.58
3	8	0.62	0.62	0.62
	32	0.65	0.57	0.61
	64	0.64	0.57	0.64
	256	0.64	0.6	0.62
	512	0.65	0.57	0.62
4	8	0.64	0.63	0.61
	32	0.64	0.67	0.61
	64	0.63	0.61	0.64
	256	0.64	0.63	0.63
	512	0.63	0.58	0.6

Table 4: نتایج عملکرد مدل GRU با تعداد مختلف ابعاد لایه مخفی و تعداد پشته‌های مختلف

# of stacks	# of hcells	Train Acc	Validation Acc	Test Acc
2	8	0.66	0.59	0.6
	32	0.64	0.62	0.61
	64	0.64	0.55	0.6
	256	0.63	0.59	0.59
	512	0.64	0.57	0.58
3	8	0.62	0.65	0.6
	32	0.64	0.63	0.61
	64	0.65	0.6	0.62
	256	0.65	0.57	0.64
	512	0.65	0.55	0.59
4	8	0.61	0.61	0.59
	32	0.63	0.63	0.63
	64	0.63	0.64	0.63
	256	0.62	0.63	0.6
	512	0.64	0.6	0.64