



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

## یادگیری عمیق

نیم سال اول ۰۱-۰۲

استاد: دکتر حمید بیگی

گردآورندگان: ایوب وزیری مقدم، محدثه میریگی و علی صفریوردی

بررسی و بازبینی: علی سلطانی

مهلت ارسال: ۱۸ آذر

RNN

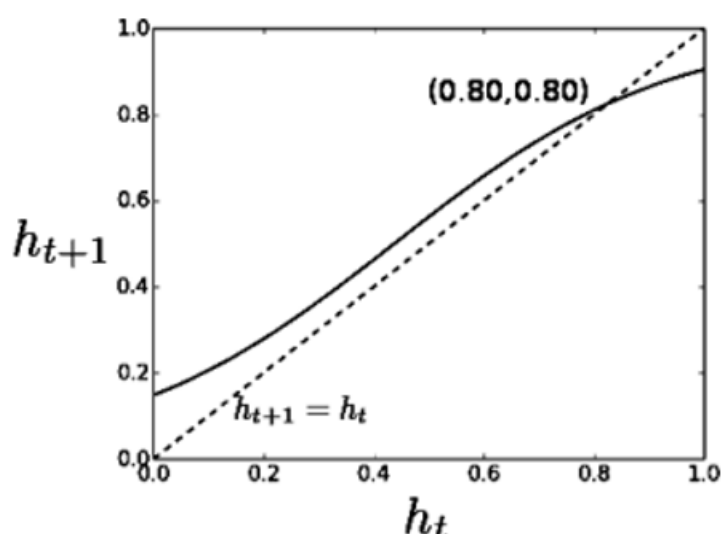
تمرین سوم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین ها بدون کسر نمره تا سقف ۱۰ روز (تا سقف ۳ روز برای هر تمرین) وجود دارد. محل بارگزاری جواب تمرین ها بعد از ۵ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ های ارسال شده پذیرفته نخواهند شد. همچنین، به ازای هر روز تأخیر غیر مجاز ۱۰ درصد از نمره تمرین به صورت ساعتی کسر خواهد شد.
- هم کاری و هم فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم فکری و یا استفاده از هر منابع خارج درسی، نام هم فکران 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.
- پاسخ تمامی سوالات (تئوری و عملی) را در یک فایل فشرده به صورت `DL_Hw3_[firstName]_[lastName]_[StudentId]` نام گذاری کرده و ارسال کنید.

سوالات نظری (۷۰ نمره)

### ۱. سوال اول (۸ نمره)

فرض کنید تابع فعالساز شبکه RNN یک سیگموید واحد است که ورودی اش همواره صفر است. شکل زیر نشان می دهد که چگونه  $h_{t+1}$  نسبت به  $h_t$  تغییر می کند. در واقع برای این شبکه  $W_2 = 3$  و  $b_2 = -1$  را در نظر بگیرید. به عبارت دیگر  $h_{t+1} = \sigma(3h_t - 1)$ . حال به سوالات زیر پاسخ دهید.



(آ) اگر طول دنباله ورودی افزایش یابد یا به عبارتی دیگر  $t \rightarrow \infty$  در  $h_{t+1}$  چه تغییری رخ می دهد؟

- (ب) برای چه مقادیری از  $h_0$  شبکه دچار محوشدگی گرادیان و انفجار گرادیان برای دنباله های طولانی می شود؟  
 (ج) نمودار plot phase برای توصیف این سیستم رسم نمایید.  
 (د) اگر تابع relu را به جای سیگموید به کار ببریم انفجار گرادیان و محوشدگی آن را بررسی کنید.

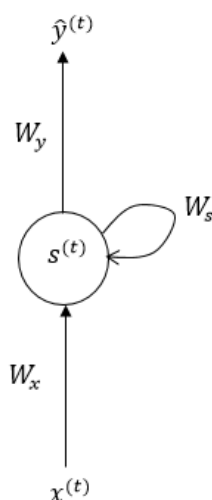
## ۲. سوال دوم (۱۲ نمره)

- (آ) فرض کنید می خواهیم ویدیوها را براساس اکشن رخ داده در آن کلاس بندی کنیم. شبکه ای مبتنی بر RNN برای کلاس بندی ویدیوها ارائه دهید.  
 (ب) یکی دیگر از کاربردهای RNN نویسه گردانی است. فرض کنید دیتاست موازی ازالفبای محلی و الفبای لاتین متناظر با آن داریم.  
 هدف این است که یک مدلی مبتنی بر RNN ارائه دهید که رشته ورودی را به کلمه ای مرتبط از زبان لاتین برگرداند.  
 i. تعداد کل محاسبات انجام شده توسط شبکه را بیابید.  
 ii. تعداد پارامترهای این شبکه را محاسبه کنید.

## ۳. الگوریتم پس انتشار در طول زمان (۱۵ نمره)

شبکه عصبی بازگشتی زیر را در نظر بگیرید که در آن  $s^{(t)}$  نشان دهنده hidden state در زمان  $t$  و  $W_s$  ماتریس وزن مرتبط با آن است،  $x^{(t)}$  نشان دهنده ورودی در زمان  $t$  و  $W_x$  ماتریس مرتبط با آن است و  $\hat{y}^{(t)}$  خروجی در زمان  $t$  و  $W_y$  ماتریس وزن مرتبط با آن است.  $\sigma$  تابع فعال ساز سیگموید و  $E^{(t)}$  تابع خطا در زمان  $t$  را نشان می دهد.

$$\begin{aligned} s^{(t)} &= W_s s^{(t-1)} + W_x x^{(t)} \\ \hat{y}^{(t)} &= \sigma(W_y s^{(t)}) \\ E^{(t)} &= (y^{(t)} - \hat{y}^{(t)})^2 \end{aligned}$$



$$(\bar{A}) \quad \frac{\partial E^{(3)}}{\partial W_x}, \frac{\partial E^{(3)}}{\partial W_s}, \frac{\partial E^{(3)}}{\partial W_y} \text{ را محاسبه کنید (با جزئیات).}$$

(ب) توضیح دهید که چرا مشکل ناپدید شدن و انفجار گرادیان در شبکه‌های عصبی بازگشتی شدیدتر است؟

#### ۴. وابستگی طولانی مدت (۱۰ نمره)

رابطه بازگشتی زیر را به عنوان یک شبکه عصبی بازگشتی بسیار ساده، بدون تابع فعال‌ساز غیرخطی و فاقد ورودی  $x$  در نظر بگیرید به طوری که در آن  $h^{(t)}$  نشان‌دهنده hidden state در زمان  $t$  و  $W$  ماتریس وزن را نشان می‌دهد.

$$h^{(t)} = W^T h^{(t-1)}$$

$$\text{اگر } W = \begin{pmatrix} 0.58 & 0.24 \\ 0.24 & 0.72 \end{pmatrix} \text{ و } h^{(0)} = I_2 \text{ (ماتریس همانی دوبعدی) باشد:}$$

(\bar{A})  $h^{(30)}$  را محاسبه کنید. (راهنمایی: از تجزیه ویژه ماتریس استفاده کنید)

(ب) در خروجی قسمت (\bar{A}) چه چیزی مشاهده می‌شود؟ چرا؟

#### ۵. بررسی اثر بلندمدت نویز در LSTM (۱۵ نمره)

یکی از موضوعاتی که در شبکه‌های یادگیری عمیق مطرح می‌شود بحث وجود نویز در داده‌ها است و اثر کوتاه مدت و طولانی مدت نویز در موارد متعددی مورد بررسی قرار گرفته است.

(\bar{A}) در مورد اثر کوتاه مدت توضیح دهید که به طور کلی در شبکه‌های عمیق چه تاثیری دارد؟

(ب) در مورد اثر بلند مدت نویزها تحقیق نمایید. در این راستا لازم است با نوشتن روابط مربوطه و ارائه تحلیل، اثر نویز را بررسی نمایید و به طور خاص بررسی نمایید نویزها آیا/چگونه در حافظه بلند مدت LSTM باقی خواهند ماند؟

(ج) آیا می‌توان روابط ارائه شده در پاسخ بخش ب را به انواع توزیع‌های نویزی بسط داد؟

#### ۶. نسخه چند بعدی شبکه LSTM (۱۰ نمره)

شبکه عصبی بازگشتی چند بعدی: مقالات مختلفی برای بهبود شبکه‌های عصبی بازگشتی، نسخه چند بعدی آن را ارائه نموده‌اند. البته توجه شود در مواردی بیان شدن عبارت چند بعدی در مقالات مختلف معانی متفاوتی دارد.

(\bar{A}) بررسی نمایید چگونه می‌توان این مفهوم را با BLSTM ترکیب نمود؟

(ب) برای این دسته از مدل‌ها برای حالت  $n$  بعدی معادله فعال‌سازی را نوشته و بررسی کنید پیچیدگی از نظر تعداد پارامترها و مدت زمان اجرا در صورت داشتن بینهایت ماشین موازی چه میزان خواهد بود؟

نکته: در صورت علاقه‌مندی به این دسته از مدل‌ها می‌توانید عبارت MDRNN را سرچ نمایید.

۱. پیاده سازی LSTM و GRU (۱۸ نمره)

در این تمرین هدف پیاده سازی دو شبکه LSTM و GRU به منظور پیش بینی تعداد مسافران از مجموعه داده مسافران خطوط هوایی (فایل Data.csv ضمیمه شده) است. در این تمرین از Adam به عنوان الگوریتم بهینه ساز و از میانگین مربعات خطا به عنوان تابع loss استفاده شود. برای حل این تمرین به نوتبوک مربوطه مراجعه کنید.

۲. سری زمانی (۱۲ نمره)

در این تمرین با داده های سری زمانی کار خواهید نمود. تلاش کنید قسمت های خالی را تکمیل نمایید و پاسخ سوالات داده شده را نیز تکمیل بفرمایید.  
توجه داشته باشید ارائه توضیح راه حل به صورت کامنت در نوتبوک مورد نیاز است.