



## یادگیری ژرف

نیم سال اول ۱۴۰۲-۰۳

مدرس: دکتر حمید بیگی

زمان تحویل: ۱۱ آذر

عنوان تمرین

تمرین سری اول (۱۰۰ نمره)

لطفا نکات زیر را رعایت کنید:

- سوالات خود را از طریق پست مربوط به تمرین در Quera مطرح کنید.
- در هر کدام از سوالات، اگر از منابع خارجی استفاده کرده‌اید باید آن را ذکر کنید. در صورت همفکری با افراد دیگر هم باید نام ایشان را در سوال مورد نظر ذکر نمایید.
- پاسخ ارسالی واضح و خوانا باشد. در غیر این صورت ممکن است منجر به از دست دادن نمره شود.
- پاسخ ارسالی باید توسط خود شما نوشته شده باشد. به اسکرین‌شات از منابع یا پاسخ افراد دیگر نمره‌ای تعلق نمی‌گیرد.
- در صورتی که بخشی از سوال‌ها را جای دیگری آپلود کرده و لینک آن را قرار داده باشید، حتما باید تاریخ آپلود مشخص و قابل اعتنا باشد.
- تمام پاسخ‌های خود را در یک فایل با فرمت HW#\_[SID]\_[Fullname].zip روی کوئرا قرار دهید.

## سوال ۱: (نظری) عنوان سوال اول (۳۰ نمره)

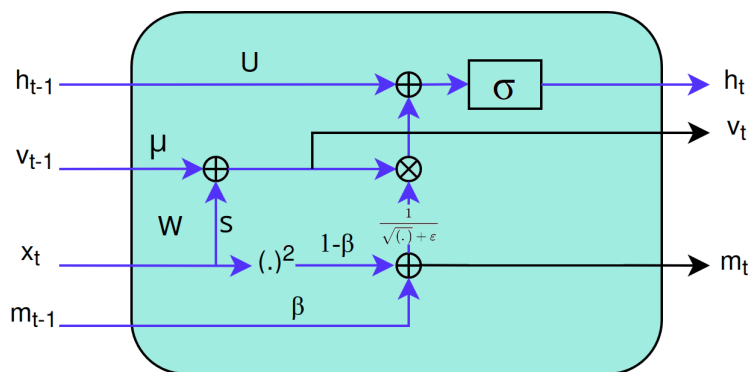
(۱) با توجه به معماری LSTM داده شده در شکل ۱ به سوالات پاسخ دهید.

(آ) معادله‌های تمام متغیرهای خروجی این سلول را بنویسید.

(ب)  $\frac{\partial h_T}{\partial x_i}$  را حساب کنید.(ج)  $\frac{\partial h_T}{\partial h_i}$  را هم حساب کنید

(د) این معماری با مشکل Gradient Vanishing یا Gradient Explosion رو به رو می‌شود؟ توضیح دهید و راه حلی برای آن ارائه دهید.

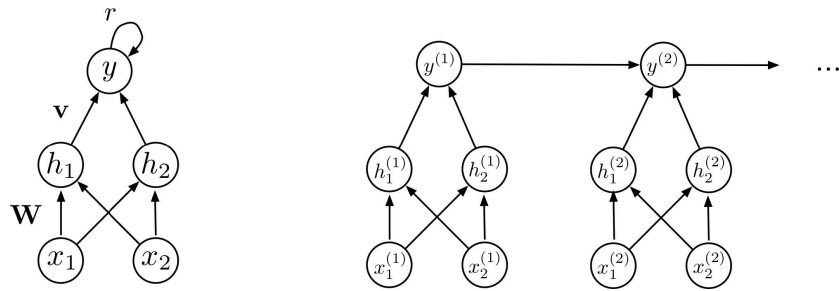
(ه) این سلول شبیه به چه الگوریتم بهینه‌سازی عمل می‌کند؟ نحوه کار این الگوریتم را توضیح دهید.



شکل ۱: LSTM

## سوال ۲: (نظری) عنوان سوال دوم (۲۰ نمره)

در شکل زیر معماری یک شبکه بازگشتی (راست) و یک سلول از آن (چپ) را مشاهده می‌کنید. می‌خواهیم این RNN را به نحوی طراحی کنیم که با دریافت دو دنباله دودویی ورودی  $\mathbf{x}_1 = (x_1^{(1)}, \dots, x_1^{(N)})$  و  $\mathbf{x}_2 = (x_2^{(1)}, \dots, x_2^{(N)})$  مشخص کند که آیا آن‌ها یکسان هستند یا نه.



در این شبکه، محاسباتی که در هر گام زمانی انجام می‌شود به صورت زیر است:

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

$$y^{(t)} = \begin{cases} \phi(v^T h^{(t)} + r y^{(t-1)} + c) & t > 1 \\ \phi(v^T h^{(t)} + c_0) & t = 1 \end{cases}$$

که در آن  $\phi$  تابع فعال‌سازی پله،  $W$  ماتریس  $2 \times 2$ ،  $b$  و بردار  $2$  بعدی،  $r$ ،  $c$  و  $c_0$  اسکالر می‌باشند. در این شبکه در هر گام زمانی  $t$ ، عناصر  $x_1^{(t)}$  و  $x_2^{(t)}$  به عنوان ورودی به مدل داده می‌شوند و خروجی  $y^{(t)}$  محاسبه می‌کند که آیا همه جفت عناصر تا زمان  $t$  مطابقت داشته‌اند یا خیر. به این ترتیب خروجی  $y^{(N)}$  نشان‌دهنده برابری / نابرابری دو دنباله دودویی ورودی خواهد بود. پارامترهای  $W$ ،  $b$ ،  $v$ ،  $r$ ،  $c$  و  $c_0$  را به نحوی تعیین کنید که شبکه داده شده تابع مورد نظر را پیاده‌سازی کند.

### سوال ۳: (عملی) تکمیل نوت بوک‌ها (۵۰ نمره)

توضیحات لازم در مورد هر نوت بوک داخل آن وجود دارد. لطفاً آن‌ها را تکمیل کنید.