به نام خدا

تمرین سری پنجم

درس یادگیری عمیق

استاد مرضیه داودآبادی

۹۹۵۲۱۴۶۹

سینا علی‌نژاد

سوال ۱-

الف) گزینه ۱ درست نیست زیرا many to many می‌باشد. گزینه دوم درست است زیرا ورودی یک متن است و خروجی تنها یک عدد است. گزینه سوم درست است زیرا گفتار یک ورودی چندتایی است و تعیین جنسیت یک خروجی تکی است.

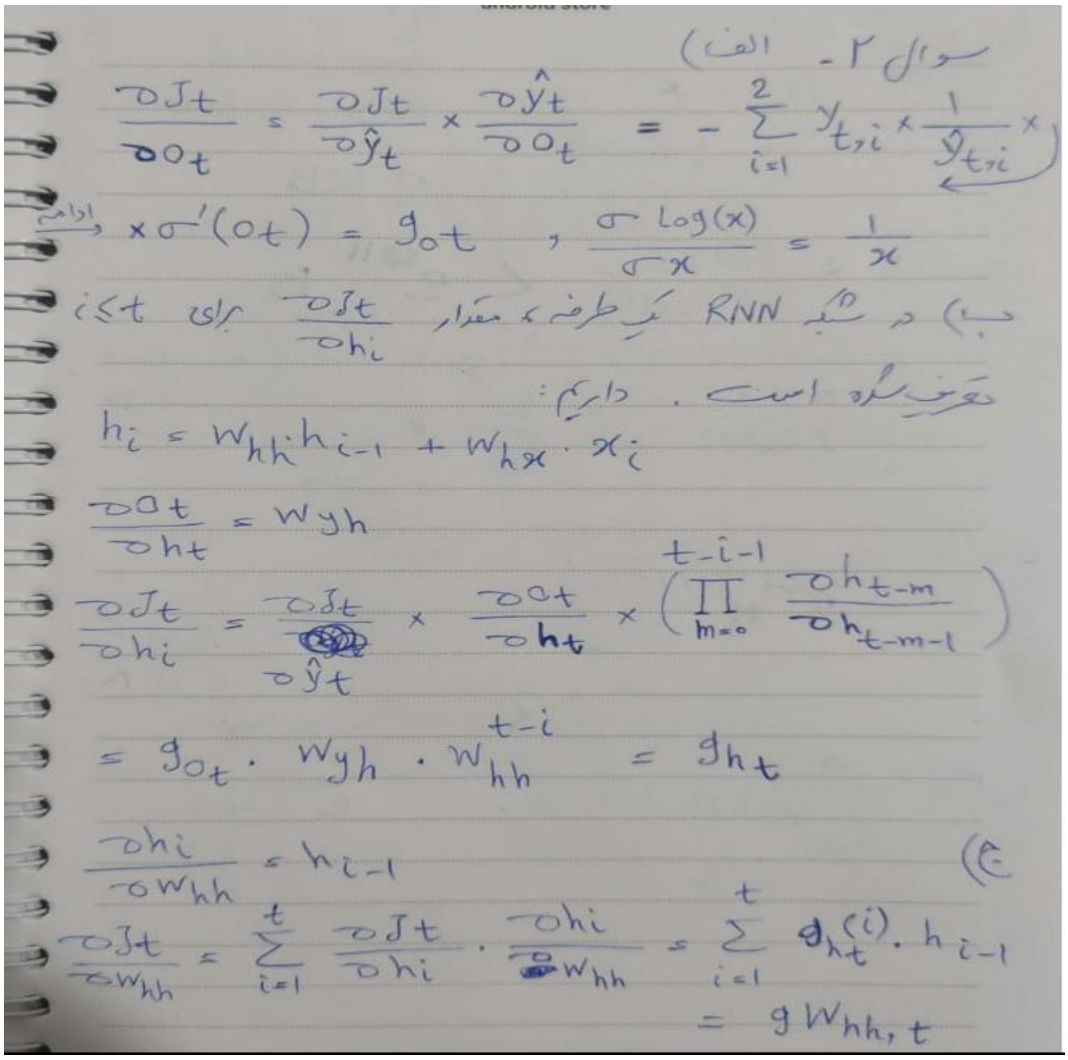
ب) مورد سوم درست است، زیرا در صورت سوال گفته شده که اخلاق گربه به هوای فعلی و چند روز گذشته وابسته است، در نتیجه نیازی به دوطرفه بودن شبکه بازگشتی نداریم و همان یک طرفه صحیح است. مورد چهارم نیز اشتباه است زیرا روزهای گذشته را در نظر نگرفته که در این صورت اصلا به شبکه بازگشتی یا RNN نیازی نخواهیم داشت.

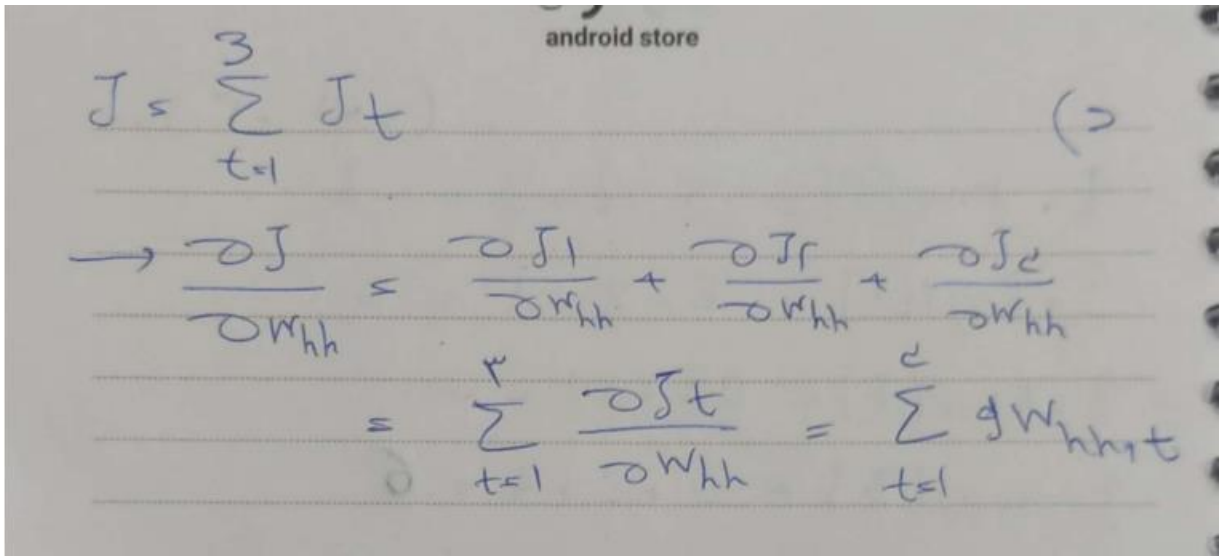
ج) مورد اول و دوم به طور واضحی اشتباه هستند پس باید از بین مورد سوم و چهارم انتخاب کنیم. مورد چهارم کاملتر است زیرا در تخمین y_t از ورودی x_t نیز استفاده می‌شود و همچنین خروجی‌های قبلی یعنی $y_{t-1}, y_{t-2}, \dots$ اما در گزینه سوم، ورودی در همان لحظه را در نظر نگرفته است پس نمیتواند درست باشد. البته در گزینه چهارم به نظرم یک اشتباه تایپی وجود دارد و باید بدین شکل باشد:

$$P(\{y_t | y_{t-1}, y_{t-2}, \dots, x_t\})$$

اگر گزینه چهارم اشتباه تایپی نباشد، گزینه سوم کاملتر می‌باشد.

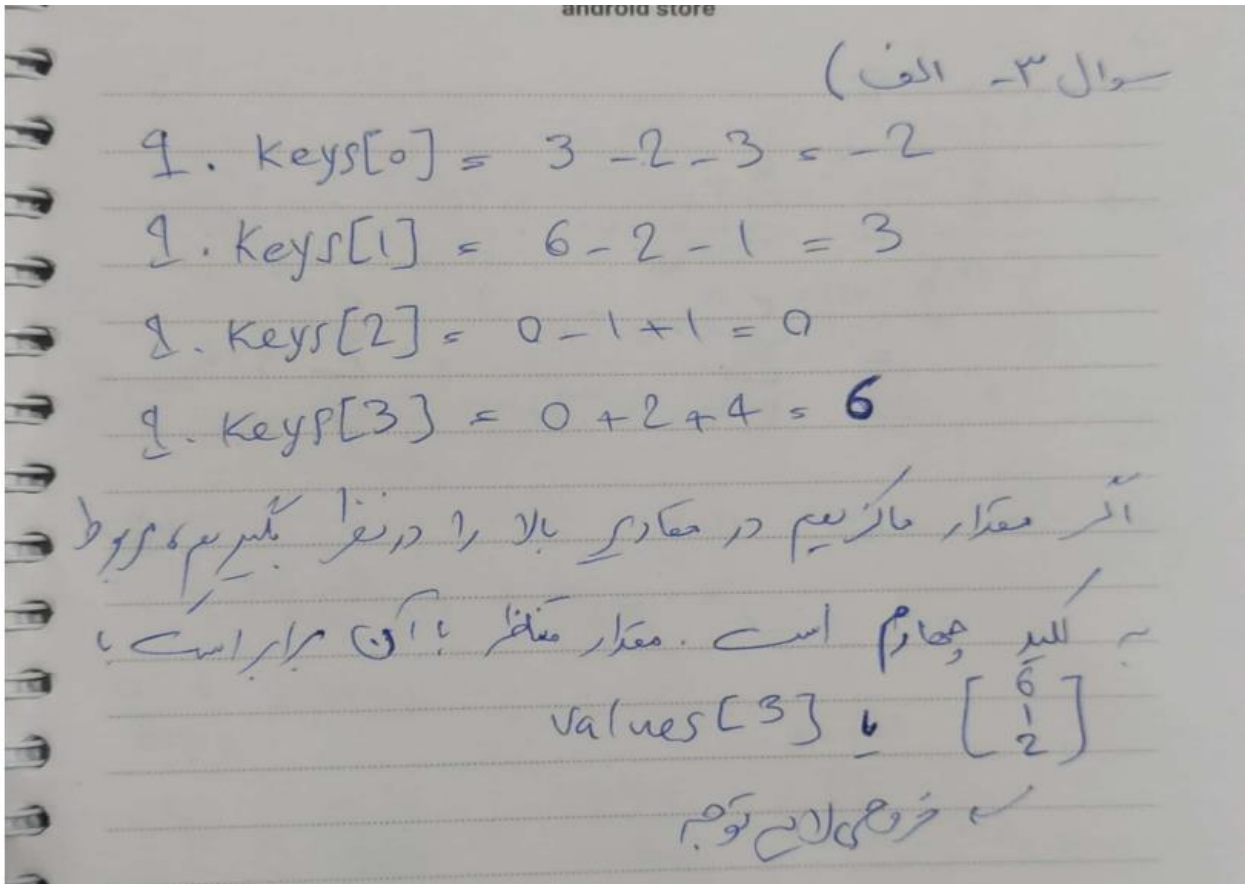
سوال ۲-


سوال ۲ - الف)
$\frac{\partial J_t}{\partial o_t} = \frac{\partial J_t}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial o_t} = -\sum_{i=1}^{2} y_{t,i} \times \frac{1}{\hat{y}_{t,i}} \times$
ادامه $\times \sigma'(o_t) = g_{o_t}$, $\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$
ب) در شبکه RNN یک طرفه، مقدار $\frac{\partial J_t}{\partial h_i}$ برای $i \leq t$ به صورت بازگشتی تعریف شده است. داریم:
$h_i = w_{hh} h_{i-1} + w_{hx} \cdot x_i$
$\frac{\partial o_t}{\partial h_t} = w_{yh}$
$\frac{\partial J_t}{\partial h_i} = \frac{\partial J_t}{\partial \hat{y}_t} \times \frac{\partial o_t}{\partial h_t} \times \left(\prod_{m=0}^{t-i-1} \frac{\partial h_{t-m}}{\partial h_{t-m-1}} \right)$
$= g_{o_t} \cdot w_{yh} \cdot w_{hh}^{t-i} = g_{h_t}$
ج)
$\frac{\partial h_i}{\partial w_{hh}} = h_{i-1}$
$\frac{\partial J_t}{\partial w_{hh}} = \sum_{i=1}^{t} \frac{\partial J_t}{\partial h_i} \cdot \frac{\partial h_i}{\partial w_{hh}} = \sum_{i=1}^{t} g_{h_t}^{(i)} \cdot h_{i-1}$
$= g_{W_{hh},t}$


android store
$J = \sum_{t=1}^{3} J_t$
$\rightarrow \frac{\partial J}{\partial w_{hh}} = \frac{\partial J_1}{\partial w_{hh}} + \frac{\partial J_2}{\partial w_{hh}} + \frac{\partial J_3}{\partial w_{hh}}$
$= \sum_{t=1}^{3} \frac{\partial J_t}{\partial w_{hh}} = \sum_{t=1}^{3} g_{W_{hh},t}$

سوال ۳-

الف)



ب)

- مشکل اول این است که تابع argmax مشتق پذیر نیست و این کار را برای اجرای الگوریتم back propagation در الگوریتم گرادیان کاهشی مشکل میکند و یادگیری را دچار اختلال میکند. حتی اگر به نحوی گرادیان را برای این تابع تعریف کنیم، این مقادیر برای نقاط به غیر از نقطه انتخاب شده، بسیار کوچک است و فرآیند یادگیری شبکه بسیار کند خواهد بود.

- اصطلاحا به استفاده از توابعی مثل argmax در مکانیزم توجه، hard attention گفته می‌شود، زیرا مدل به هیچ نقطه دیگری به غیر از شبیه ترین نقطه توجه نمیکند، در برخی کاربردها مثل machine translation لازم است که مدل برای ترجمه به چندین بخش زبان مبدا دقت کند. یا ممکن است نقطه‌ای که بیشترین شباهت را داشته، یک نقطه نویزی باشد و نیاز باشد به نقاط اطراف هم توجه کنیم.

برای حل این مشکل میتوان از تابع softmax برای مکانیزم توجه استفاده کرد، در این صورت مقادیر توجه بین همه پخش میشود ولی نقطه با بیشترین شباهت ضریب بزرگتری نسبت به بقیه میگیرد.