

(Q1)

(a)

برای گزینه آ مناسب نیست. چون در تشخیص گفتار، ورودی یک دنباله صوتی از کلمات و خروجی هم یک دنباله نوشتاری از کلمات است و به همین دلیل مدل many to one مناسب این کار نیست.

برای گزینه ب مناسب است. چون ورودی یک دنباله متنی است و خروجی هم تنها یک نوروں برای نشان دادن مثبت یا منفی بودن نتیجه نیاز دارد.

برای گزینه ج هم مناسب است. چون باز هم مثل گزینه ب ورودی یک دنباله و خروجی هم تنها یک نوروں نیاز دارد. برای نشان دادن مرد یا زن بودن گوینده.

(b)

گزینه ج صحیح است. از شبکه یک طرفه استفاده میکنم چون خروجی در هر مرحله، تنها به مراحل قبلی بستگی دارد. یعنی گره از وضعیت هوا در روزهای آینده خبر ندارد پس اخلاقی هم تأثیری از داده های روز های آینده نمیپذیرد.

(c)

گزینه ج صحیح است.

(Q2)

$$J_t = - \sum_{i=1}^2 y_{t,i} \log \hat{y}_{t,i}$$

(الف)

$$\frac{\partial J_t}{\partial \theta_i} = \frac{\partial J_t}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial \theta_i} = \sigma'(\theta_i) \times \begin{bmatrix} -\frac{y_{t,1}}{\hat{y}_{t,1}} & -\frac{y_{t,2}}{\hat{y}_{t,2}} \end{bmatrix} = \sigma'(\theta_i) \times \frac{-y_t}{\hat{y}_t}$$

$$\frac{\partial J_t}{\partial h_i} \text{ , } i=3 = \frac{\partial J_t}{\partial \theta_3} \times \frac{\partial \theta_3}{\partial h_3} = g_{\theta,3} \times w_{yh,3} \quad \leftarrow$$

$$\frac{\partial J_t}{\partial w_{hh}} = \sum_{i=1}^I \frac{\partial J_t}{\partial h_i} \times \frac{\partial h_i}{\partial w_{hh}} = \sum_{i=1}^I g_{h,i} \times h_{i-1} \quad (2)$$

$$\frac{\partial J}{\partial w_{hh}} = \frac{\partial}{\partial w_{hh}} \left(\sum_{t=1}^3 J_t \right) = \sum_{t=1}^3 \frac{\partial J_t}{\partial w_{hh}} = \sum_{t=1}^3 g_{w_{hh},t} \quad (3)$$

(Q3)

(الف)

میزان شباهت برای کلید های مختلف:

K4:6 K3:0 K2:3 K1:-2

بنابراین کلید چهارم انتخاب میشود. و خروجی برابر با [6 1 2] خواهد بود.

(ب)

استفاده از argmax به عنوان مکانیزم توجه میتواند یادگیری را برای مدل سخت تر کند. چون نمیتوان از این تابع گرادیان گرفت. همچنین در صورت ایجاد گرادیان جزعی، اندازه گرادیان بسیار کوچک خواهد بود.

از طرفی استفاده از این مکانیزم باعث میشود که مدل به سایر value ها هیچ توجهی نکند و تنها بر اساس یکی از value ها خروجی و هزینه را محاسبه کند. استفاده از توابعی که توجه مدل را روی تمام نمونه ها توزیع کند میتواند نتایج بهتری داشته باشد. مثل تابع softmax .