

①

الف)

z_t : باقی‌مانده رابطه زیر:

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

آر z_t صفر باشد h_t تأثیر ندارد خواهد بود و اطلاعات در دوره جدیداً اثر ندارد خواهند بود.
باقی‌مانده این رابطه ارزش در هر حالت قبلی را مشخص می‌کند، z_t ، بیت آپدیت است.

r_t : باقی‌مانده رابطه زیر:

$$\tilde{h}_t = \tanh(w^{(hx)} x_t + r_t w^{(hh)} h_{t-1})$$

آر r_t صفر باشد، حالت قبلی تأثیر ندارد خواهد بود. در واقع مشخص می‌کند تأثیر هر کدام از دوره‌های
حالت قبلی در شکل حالت جدید چه قدرات. پس r_t ، بیت ریست (reset) است.

$$\frac{\partial J}{\partial w} = \sum_{k=1}^t \frac{\partial J}{\partial y_k} \frac{\partial y_k}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial w}$$

نرم سیم به تدریج تأثیر ندارد در هر زمان gradient vanishing

$$\frac{\partial h_t}{\partial h_k} = \prod_{j=k+1}^t \frac{\partial h_j}{\partial h_{j-1}}$$

x_{NN} در رابطه: $\frac{\partial h_j}{\partial h_{j-1}} = z_j + (1 - z_j) \frac{\partial \tilde{h}_j}{\partial h_{j-1}}$

آر $z_j = 1 \Rightarrow$

عادل ضرب همگرا متوالی تبدیل به $\frac{\partial h_j}{\partial h_{j-1}} = 1$

بیت ریست (reset) gradient vanishing