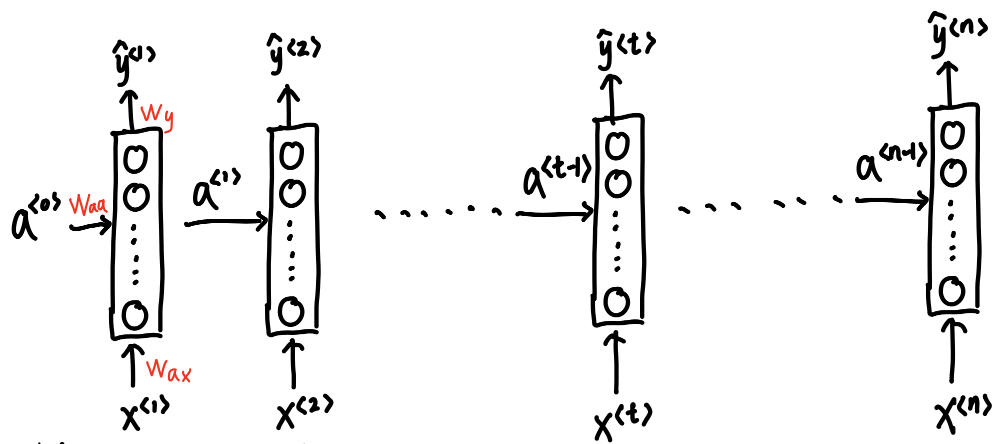
一. RNN

序列： I will become ···· the ····· best

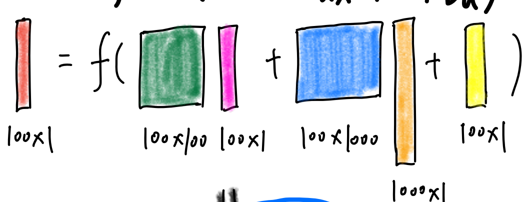
$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ ··· $x^{<t>}$ ··· $x^{<n>}$



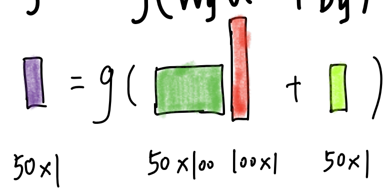
$a^{<1>}=f(W_{aa}a^{<0>}+W_{ax}x^{<1>}+b_a)$ $a^{<2>}=f(W_{aa}a^{<1>}+W_{ax}x^{<2>}+b_a)$ $a^{<t>}=f(W_{aa}a^{<t-1>}+W_{ax}x^{<t>}+b_a)$ $a^{<n>}=f(W_{aa}a^{<n-1>}+W_{ax}x^{<n>}+b_a)$

$\hat{y}^{<1>}=g(W_y a^{<1>}+b_y)$ $\hat{y}^{<2>}=g(W_y a^{<2>}+b_y)$ $\hat{y}^{<t>}=g(W_y a^{<t>}+b_y)$ $\hat{y}^{<n>}=g(W_y a^{<n>}+b_y)$

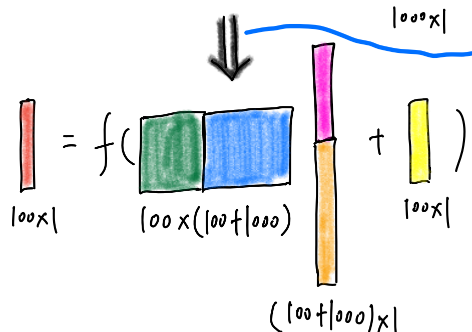
$$a^{<t>}=f(W_{aa}a^{<t-1>}+W_{ax}x^{<t>}+b_a)$$



$$\hat{y}^{<t>}=g(W_y a^{<t>}+b_y)$$



分块矩阵性质：$[A,B]\begin{bmatrix}C\\D\end{bmatrix}=AC+BD$



$$a^{<t>}=f\left([W_{aa}|W_{ax}]\begin{bmatrix}a^{<t-1>}\\x^{<t>}\end{bmatrix}+b_a\right) \Rightarrow a^{<t>}=f\left(W_a\begin{bmatrix}a^{<t-1>}\\x^{<t>}\end{bmatrix}+b_a\right)$$

每一层的神经元可以看作 Wa 的行向量

Wy 不能具象到神经元中

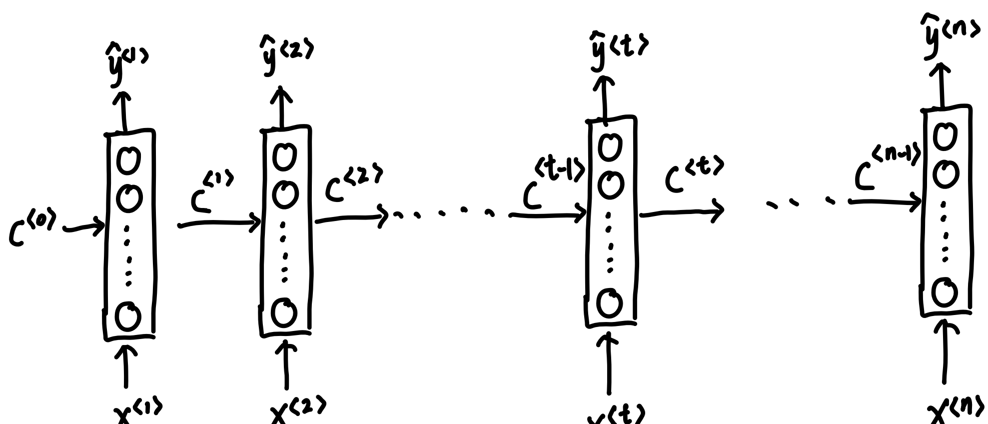
$$L^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -y^{<t>}\log\hat{y}^{<t>} - (1-y^{<t>})\log(1-\hat{y}^{<t>})$$

$$J(W_a, W_y, b_a, b_y) = L(\hat{y}, y) = \sum_{t=1}^{n} L^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

二、GRU

序列： I will become …… the …… best

$\downarrow$ $x^{<1>}$ $x^{<2>}$ $x^{<3>}$ --- $x^{<t>}$ --- $x^{<n>}$



$\Gamma_u = \sigma(W_u \begin{bmatrix} c^{<0>} \\ x^{<1>} \end{bmatrix} + b_u)$

$\Gamma_r = \sigma(W_r \begin{bmatrix} c^{<0>} \\ x^{<1>} \end{bmatrix} + b_r)$

$\tilde{c}^{<1>} = f(W_c \begin{bmatrix} \Gamma_r * c^{<0>} \\ x^{<1>} \end{bmatrix} + b_c)$

$c^{<1>} = \Gamma_u * \tilde{c}^{<1>} + (1-\Gamma_u)c^{<0>}$

$\hat{y}^{<1>} = g(W_y c^{<1>} + b_y)$

$\Gamma_u = \sigma(W_u \begin{bmatrix} c^{<1>} \\ x^{<2>} \end{bmatrix} + b_u)$

$\Gamma_r = \sigma(W_r \begin{bmatrix} c^{<1>} \\ x^{<2>} \end{bmatrix} + b_r)$

$\tilde{c}^{<2>} = f(W_c \begin{bmatrix} \Gamma_r * c^{<1>} \\ x^{<2>} \end{bmatrix} + b_c)$

$c^{<2>} = \Gamma_u * \tilde{c}^{<2>} + (1-\Gamma_u)c^{<1>}$

$\hat{y}^{<2>} = g(W_y c^{<2>} + b_y)$

$\Gamma_u = \sigma(W_u \begin{bmatrix} c^{<t-1>} \\ x^{<t>} \end{bmatrix} + b_u)$

$\Gamma_r = \sigma(W_r \begin{bmatrix} c^{<t-1>} \\ x^{<t>} \end{bmatrix} + b_r)$

$\tilde{c}^{<t>} = f(W_c \begin{bmatrix} \Gamma_r * c^{<t-1>} \\ x^{<t>} \end{bmatrix} + b_c)$

$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1-\Gamma_u)c^{<t-1>}$

$\hat{y}^{<t>} = g(W_y c^{<t>} + b_y)$

$\Gamma_u = \sigma(W_u \begin{bmatrix} c^{<n-1>} \\ x^{<n>} \end{bmatrix} + b_u)$

$\Gamma_r = \sigma(W_r \begin{bmatrix} c^{<n-1>} \\ x^{<n>} \end{bmatrix} + b_r)$

$\tilde{c}^{<n>} = f(W_c \begin{bmatrix} \Gamma_r * c^{<n-1>} \\ x^{<n>} \end{bmatrix} + b_c)$

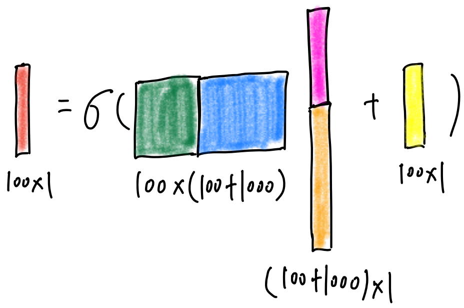
$c^{<n>} = \Gamma_u * \tilde{c}^{<n>} + (1-\Gamma_u)c^{<n-1>}$

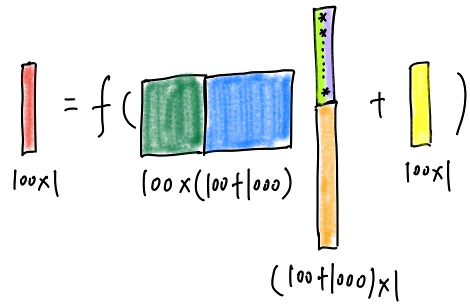
$\hat{y}^{<n>} = g(W_y c^{<n>} + b_y)$

$$\Gamma_u = \sigma\left(W_u \begin{bmatrix} c^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_u\right)$$

$$\tilde{c}^{\langle t\rangle} = f\left(W_c \begin{bmatrix} \Gamma_r * c^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_c\right)$$

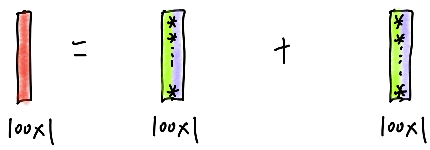
$$\Gamma_r = \sigma\left(W_r \begin{bmatrix} c^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_r\right)$$

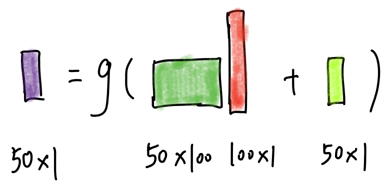




$$c^{\langle t\rangle} = \Gamma_u * \tilde{c}^{\langle t\rangle} + (1-\Gamma_u)\, c^{\langle t-1\rangle}$$

$$\hat{y}^{\langle t\rangle} = g(W_y c^{\langle t\rangle} + b_y)$$





update更新门 Γ_u ：更新多少信息.

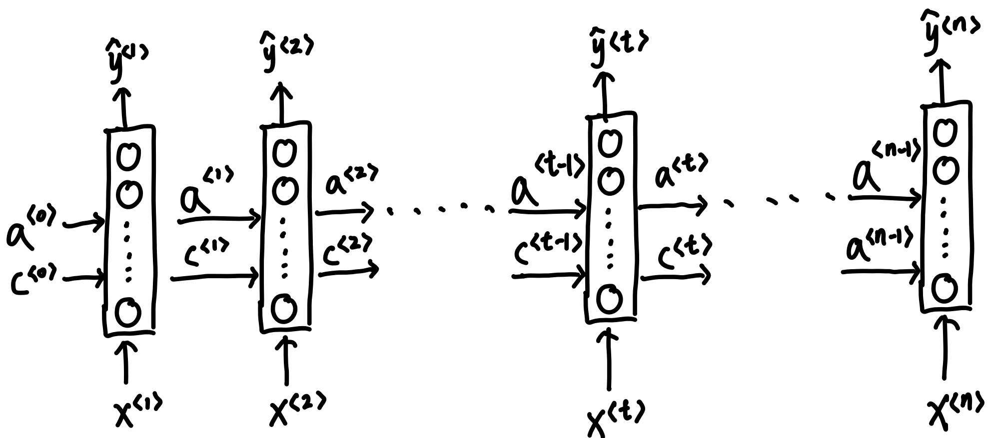
reset 重置门 Γ_r ：上一时刻信息和本时刻输入信息的权重

Γ_u和Γ_r 本质上为每个元素介于 [0,1] 之间的向量，维度与 $c^{\langle t\rangle}$(或与每层神经元个数)相等

三. LSTM

序列： I will become ···· the ···· best

$x^{\langle 1\rangle}$ $x^{\langle 2\rangle}$ $x^{\langle 3\rangle}$ --- $x^{\langle t\rangle}$ --- $x^{\langle n\rangle}$



$$\hat{c}^{\langle 1\rangle} = g(W_c \begin{bmatrix} a^{\langle 0\rangle} \\ x^{\langle 1\rangle} \end{bmatrix} + b_c)$$
$$\Gamma_u = \sigma(W_u \begin{bmatrix} a^{\langle 0\rangle} \\ x^{\langle 1\rangle} \end{bmatrix} + b_u)$$
$$\Gamma_f = \sigma(W_f \begin{bmatrix} a^{\langle 0\rangle} \\ x^{\langle 1\rangle} \end{bmatrix} + b_f)$$
$$\Gamma_o = \sigma(W_o \begin{bmatrix} a^{\langle 0\rangle} \\ x^{\langle 1\rangle} \end{bmatrix} + b_o)$$
$$c^{\langle 1\rangle} = \Gamma_u * \hat{c}^{\langle 1\rangle} + \Gamma_f * c^{\langle 0\rangle}$$
$$a^{\langle 1\rangle} = \Gamma_o * g(c^{\langle 1\rangle})$$
$$\hat{y}^{\langle 1\rangle} = g(W_y a^{\langle 1\rangle} + b_y)$$

$$\hat{c}^{\langle 2\rangle} = g(W_c \begin{bmatrix} a^{\langle 1\rangle} \\ x^{\langle 2\rangle} \end{bmatrix} + b_c)$$
$$\Gamma_u = \sigma(W_u \begin{bmatrix} a^{\langle 1\rangle} \\ x^{\langle 2\rangle} \end{bmatrix} + b_u)$$
$$\Gamma_f = \sigma(W_f \begin{bmatrix} a^{\langle 1\rangle} \\ x^{\langle 2\rangle} \end{bmatrix} + b_f)$$
$$\Gamma_o = \sigma(W_o \begin{bmatrix} a^{\langle 1\rangle} \\ x^{\langle 2\rangle} \end{bmatrix} + b_o)$$
$$c^{\langle 2\rangle} = \Gamma_u * \hat{c}^{\langle 2\rangle} + \Gamma_f * c^{\langle 1\rangle}$$
$$a^{\langle 2\rangle} = \Gamma_o * g(c^{\langle 2\rangle})$$
$$\hat{y}^{\langle 2\rangle} = g(W_y a^{\langle 2\rangle} + b_y)$$

$$\hat{c}^{\langle t\rangle} = g(W_c \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_c)$$
$$\Gamma_u = \sigma(W_u \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_u)$$
$$\Gamma_f = \sigma(W_f \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_f)$$
$$\Gamma_o = \sigma(W_o \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_o)$$
$$c^{\langle t\rangle} = \Gamma_u * \hat{c}^{\langle t\rangle} + \Gamma_f * c^{\langle t-1\rangle}$$
$$a^{\langle t\rangle} = \Gamma_o * g(c^{\langle t\rangle})$$
$$\hat{y}^{\langle t\rangle} = g(W_y a^{\langle t\rangle} + b_y)$$

$$\hat{c}^{\langle n\rangle} = g(W_c \begin{bmatrix} a^{\langle n-1\rangle} \\ x^{\langle n\rangle} \end{bmatrix} + b_c)$$
$$\Gamma_u = \sigma(W_u \begin{bmatrix} a^{\langle n-1\rangle} \\ x^{\langle n\rangle} \end{bmatrix} + b_u)$$
$$\Gamma_f = \sigma(W_f \begin{bmatrix} a^{\langle n-1\rangle} \\ x^{\langle n\rangle} \end{bmatrix} + b_f)$$
$$\Gamma_o = \sigma(W_o \begin{bmatrix} a^{\langle n-1\rangle} \\ x^{\langle n\rangle} \end{bmatrix} + b_o)$$
$$c^{\langle n\rangle} = \Gamma_u * \hat{c}^{\langle n\rangle} + \Gamma_f * c^{\langle n-1\rangle}$$
$$a^{\langle n\rangle} = \Gamma_o * g(c^{\langle n\rangle})$$
$$\hat{y}^{\langle n\rangle} = g(W_y a^{\langle n\rangle} + b_y)$$

update 更新门：加入多入 $\hat{c}^{\langle t\rangle}$ 的信息

forget 遗忘门：保留多少 $c^{\langle t-1\rangle}$ 的信息

output 输出门：输出多少 $c^{\langle t\rangle}$ 的信息

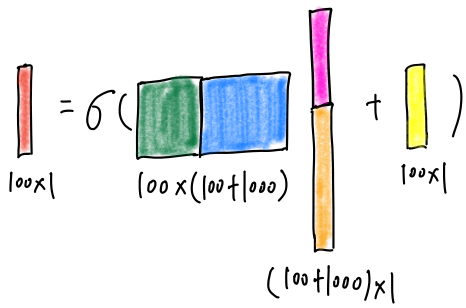
$$\hat{c}^{\langle t\rangle} = g\left(W_c \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_c\right)$$

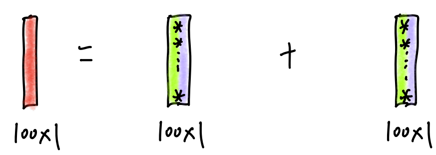
$$\Gamma_u = \sigma\left(W_u \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_u\right)$$

$$\Gamma_f = \sigma\left(W_f \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_f\right)$$

$$\Gamma_o = \sigma\left(W_o \begin{bmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{bmatrix} + b_o\right)$$

$$c^{\langle t\rangle} = \Gamma_u * \hat{c}^{\langle t\rangle} + \Gamma_f * c^{\langle t-1\rangle}$$





$$a^{\langle t\rangle} = \Gamma_o * g(c^{\langle t\rangle})$$

$$\hat{y}^{\langle t\rangle} = g(W_y a^{\langle t\rangle} + b_y)$$