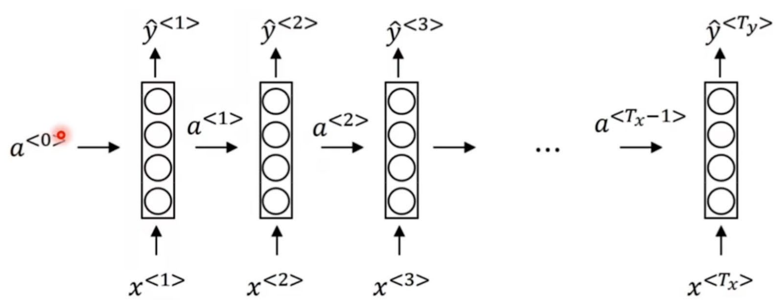
RNN

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

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

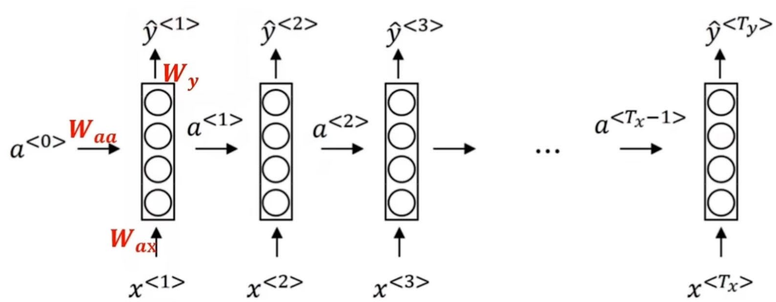


一般这个a^0都会设置为零向量

RNN

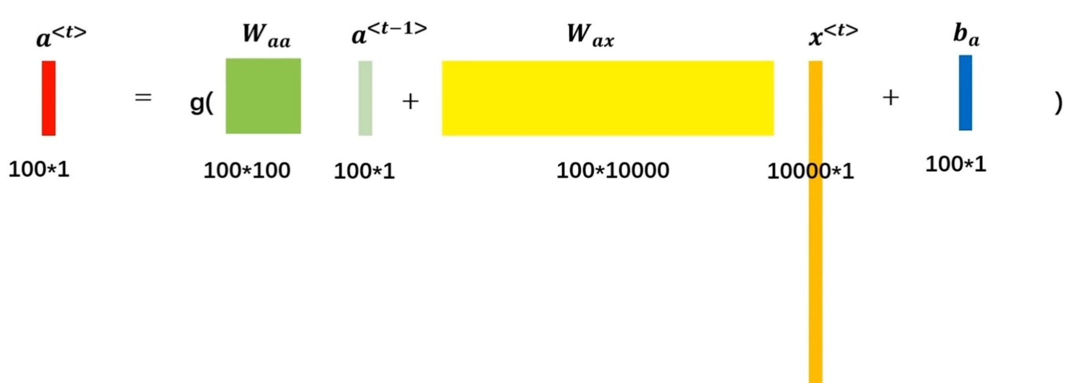
$$a^{<t>} = g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

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



RNN

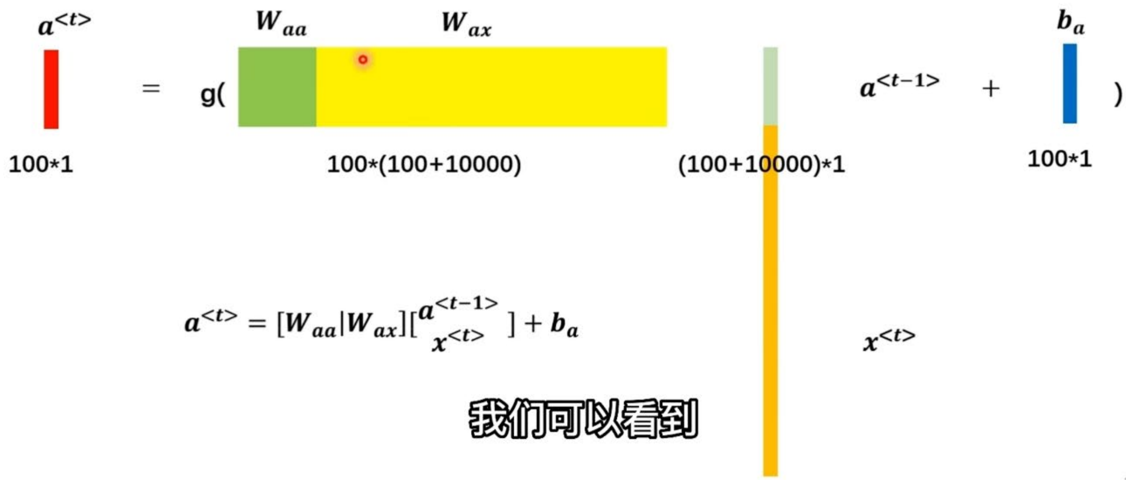
$$a^{<t>} = g(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$



经过一个激活函数g后得到

RNN

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

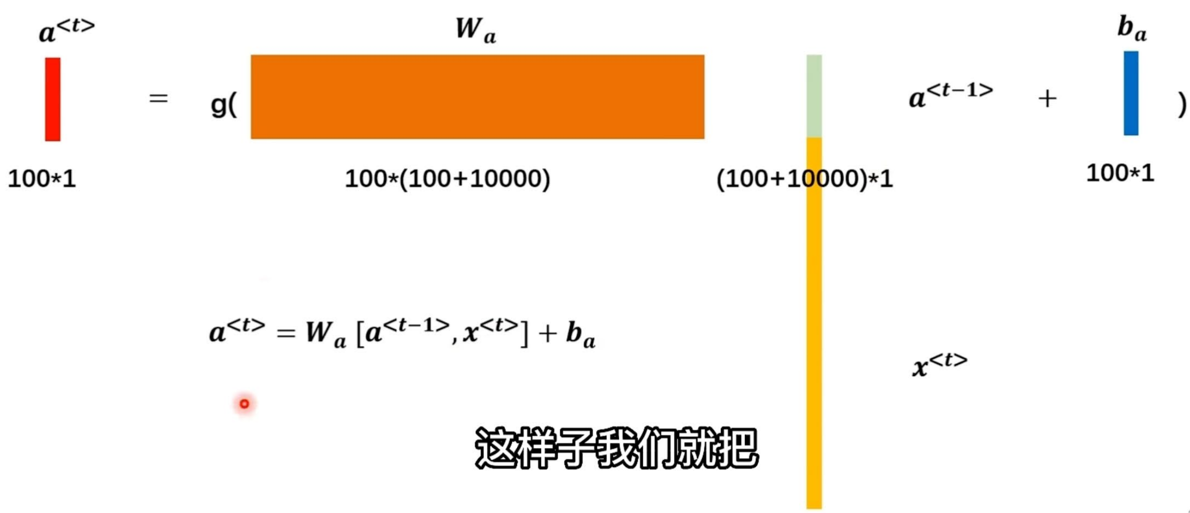


$$a^{<t>} = [W_{aa}|W_{ax}]\begin{bmatrix} a^{<t-1>} \\ x^{<t>} \end{bmatrix} + b_a$$

我们可以看到

RNN

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

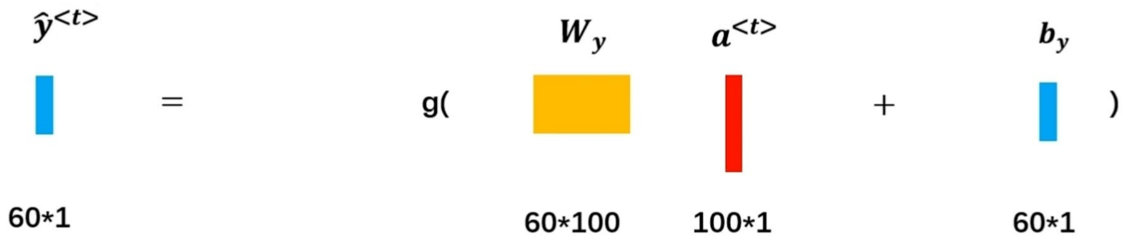


$$a^{<t>} = W_a\,[a^{<t-1>}, x^{<t>}] + b_a$$

这样子我们就把

RNN

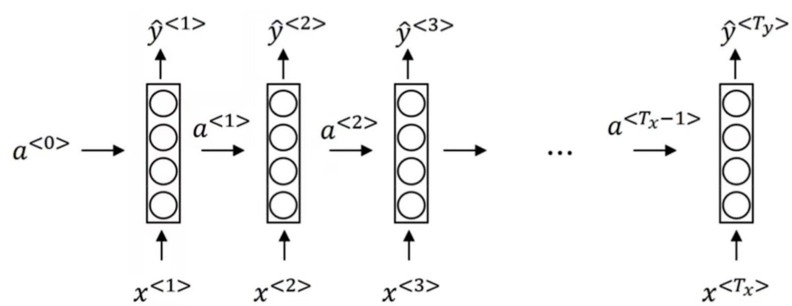
$$\hat{y}^{<t>} = g(W_y a^{<t>} + b_y)$$



RNN

$$a^{<t>} = g(W_a[a^{<t-1>}, x^{<t>}] + b_a)$$

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



GRU单元

$$a^{<t>} = c^{<t>}$$

$$c^{<t>} = g(W_c[c^{<t-1>}, x^{<t>}] + b_c)$$

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

$c^{<t>}$ 记忆单元Memory Cell

更新门： 更新多少信息

重置门： 上一时段信息和输入信息的权重

GRU单元
更新门介绍

$a^{<t>} = c^{<t>}$

$\hat{c}^{<t>} = g(W_c\,[c^{<t-1>}, x^{<t>}] + b_c)$　　完全转换：$\hat{c}^{<t>}$　完全没变：$c^{<t-1>}$　转换多少：Γ_u

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

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

$c^{<t>}$ 记忆单元Memory Cell

因为gamma_u介于0到1之间

GRU单元
重置门介绍

$a^{<t>} = c^{<t>}$

$\hat{c}^{<t>} = g(W_c\,[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$

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

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

$c^{<t>}$ 记忆单元Memory Cell

$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$

$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$

然后呢经过一个sigmoid的激活函数

GRU单元

$$\hat{c}^{<t>} = g(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

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

$$a^{<t>} = c^{<t>}$$

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

LSTM

LSTM公式

$$a^{<t>} \neq c^{<t>}$$

$$\boxed{\begin{aligned} \hat{c}^{<t>} &= g(W_c[a^{<t-1>}, x^{<t>}] + b_c) \\ \Gamma_u &= \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \\ \Gamma_f &= \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \\ \Gamma_o &= \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \end{aligned}}$$

Update更新门：加入多少$\hat{c}^{<t>}$的信息

Forget 遗忘门：保留多少$c^{<t-1>}$的信息

Output输出门：输出多少$c^{<t>}$的信息

$$c^{<t>} = \Gamma_u * \hat{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * g(c^{<t>})$$

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