Recurrent Neural Network Algorithms

Ramkumar

April 26, 2025

Algorithm RNN training

---

load dataset $\{x_t, y_t\},\ t = 1, 2, \ldots, T$

Perform encoding on $\{x_t\}$ ▷ word embedding

Choose $N^{(1)}$ ▷ hidden neurons count

Choose α ▷ learning rate

initialize $a_0 = \{0\}$, $\frac{\partial a_0}{\partial W} = \{0\}$ ▷ initializing 0^{th} hidden vector/tensor values

initialize $\frac{\partial a_0}{\partial U} = \{0\}$, $\frac{\partial a_0}{\partial b} = \{0\}$

randomly initialize W,U,b,V and c ▷ initializing model parameters

repeat

  Call Forward Propagation ( )

  Call BPTT ( )

  Call Parameter Update ( )

until Convergence

---

function Forward Propagation ( )

  for $t = 1, 2, \ldots, T$ **do**

    $z_t = Wa_{t-1} + Ux_t + b$ ▷ recurrent layer equations

    $a_t = h(z_t)$

    $o_t = Va_t + c$ ▷ output layer equations

    $\hat{y}_t = softmax(o_t)$

  end for

  Perform decoding on $\{\hat{y}_t\}$ ▷ For validation

end function

---

function BPTT( ) $\triangleright$ Back Propagation Through Time

for $t = 1, 2, \ldots, T$ **do**

Compute the following in the order

$$\delta_t^{(2)} = diag(\hat{\mathbf{y}}_t - \mathbf{y}_t)$$

$$\frac{\partial L_t}{\partial \mathbf{c}} = \delta_t^{(2)}$$

$$\frac{\partial \mathbf{o}_t}{\partial V} = reshape\left(\mathbf{a}_t^T \otimes \mathbf{I}_{N^{(2)}},\ N^{(2)} \times N^{(2)} \times N^{(1)}\right)$$

$$\frac{\partial L_t}{\partial V} = reshape\left(\delta_t^{(2)}\ reshape\left(\frac{\partial \mathbf{o}_t}{\partial \mathbf{V}},\ N^{(2)} \times \left(N^{(2)} * N^{(1)}\right)\right),\ N^{(2)} \times N^{(2)} \times N^{(1)}\right)$$

$$\frac{\partial L_t}{\partial \mathbf{a}_t} = \delta_t^{(2)}\mathbf{V}$$

$$\delta_t^{(1)} = \frac{\partial L_t}{\partial \mathbf{a}_t}\ diag(h'(\mathbf{z}_t))$$

$$\frac{\partial L_t}{\partial \mathbf{b}} = \delta_t^{(1)}\left[\mathbf{I}_{N^{(1)}} + \mathbf{W}\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{b}}\right]$$

$$\frac{\partial \mathbf{a}_t}{\partial \mathbf{b}} = diag(h'(\mathbf{z}_t))$$

$$\frac{\partial \mathbf{W}\mathbf{a}_{t-1}}{\partial \mathbf{W}} = reshape\left(\mathbf{a}_{t-1}^T \otimes \mathbf{I}_{N^{(1)}},\ N^{(1)} \times N^{(1)} \times N^{(1)}\right)$$

$$\mathbf{W}\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{W}} = reshape\left(\mathbf{W}\ reshape\left(\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{W}},\ N^{(1)} \times \left(N^{(1)} * N^{(1)}\right)\right),\ N^{(1)} \times N^{(1)} \times N^{(1)}\right)$$

$$\omega_1 = \frac{\partial \mathbf{W}\mathbf{a}_{t-1}}{\partial \mathbf{W}} + \mathbf{W}\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{W}}$$

$$\frac{\partial L_t}{\partial \mathbf{W}} = reshape\left(\delta_t^{(1)}\ reshape\left(\omega_1,\ N^{(1)} \times \left(N^{(1)} * N^{(1)}\right)\right),\ N^{(2)} \times N^{(1)} \times N^{(1)}\right)$$

$$\frac{\partial \mathbf{a}_t}{\partial \mathbf{W}} = diag(h'(\mathbf{z}_t))\omega_1$$

$$\frac{\partial \mathbf{U}\mathbf{x}_t}{\partial \mathbf{U}} = reshape\left(\mathbf{x}_t^T \otimes \mathbf{I}_{N^{(1)}},\ N^{(1)} \times N^{(1)} \times n\right)$$

$$\mathbf{W}\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{U}} = reshape\left(\mathbf{W}\ reshape\left(\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{U}},\ N^{(1)} \times \left(N^{(1)} * n\right)\right),\ N^{(1)} \times N^{(1)} \times n\right)$$

$$\omega_2 = \frac{\partial \mathbf{U}\mathbf{x}_t}{\partial \mathbf{U}} + \mathbf{W}\frac{\partial \mathbf{a}_{t-1}}{\partial \mathbf{U}}$$

$$\frac{\partial L_t}{\partial \mathbf{U}} = reshape\left(\delta_t^{(1)}\ reshape\left(\omega_2,\ N^{(1)} \times \left(N^{(1)} * n\right)\right),\ N^{(2)} \times N^{(1)} \times n\right)$$

$$\frac{\partial \mathbf{a}_t}{\partial \mathbf{U}} = diag(h'(\mathbf{z}_t))\omega_2$$

end for

end function

function PARAMETER UPDATE( )

$$\mathbf{W}_{i,j} := \mathbf{W}_{i,j} - \alpha \sum_{t=1}^{T} \sum_{k=1}^{N^{(2)}} \left(\frac{\partial L_t}{\partial \mathbf{W}} \right)_{k,i,j}, \ i = 1, 2, \ldots, N^{(1)}, \ j = 1, 2, \ldots, N^{(1)}$$

$$\mathbf{U}_{i,j} := \mathbf{U}_{i,j} - \alpha \sum_{t=1}^{T} \sum_{k=1}^{N^{(2)}} \left(\frac{\partial L_t}{\partial \mathbf{U}} \right)_{k,i,j}, \ i = 1, 2, \ldots, N^{(1)}, \ j = 1, 2, \ldots, n$$

$$\mathbf{b}_{i} := \mathbf{b}_{i} - \alpha \sum_{t=1}^{T} \sum_{k=1}^{N^{(2)}} \left(\frac{\partial L_t}{\partial \mathbf{b}} \right)_{k,i}, \ i = 1, 2, \ldots, N^{(1)}$$

$$\mathbf{V}_{i,j} := \mathbf{V}_{i,j} - \alpha \sum_{t=1}^{T} \sum_{k=1}^{N^{(2)}} \left(\frac{\partial L_t}{\partial \mathbf{V}} \right)_{k,i,j}, \ i = 1, 2, \ldots, N^{(2)}, \ j = 1, 2, \ldots, N^{(1)}$$

$$\mathbf{c}_{i} := \mathbf{c}_{i} - \alpha \sum_{t=1}^{T} \sum_{k=1}^{N^{(2)}} \left(\frac{\partial L_t}{\partial \mathbf{c}} \right)_{k,i}, \ i = 1, 2, \ldots, N^{(2)}$$

end function