Renee Senining

Info 159

Long Homework 1 Writeup

1 Preliminaries

1. $\frac{dcx}{dx} = c$
2. $\frac{d\log(x)}{dx} = 1/x$
3. $\frac{d\sigma(x)}{dx} = (\frac{1}{1+e^{-x}})' = (\frac{1}{(1+e^{-x})^2})(e^{-x}) = \sigma(x)(1-\sigma(x))$
4. $\frac{d\tanh(x)}{dx} = \frac{d}{dx}(\frac{1-e^{-2x}}{1+e^{-2x}}) = -(\tanh(x)-1)(\tanh(x)+1) = 1 = \tanh^2(x)$
5. $\frac{d(2x)^2}{dx} = 8x$

2 Forward

3 Backward

3.1 Gradient

$\frac{\partial A}{\partial U} = \frac{\partial y\log(\sigma(Vh))}{\partial\sigma(Vh)} \times \frac{\partial\sigma(Vh)}{\partial(U)}$

$\frac{\partial B}{\partial U} = \frac{\partial(1-y)\log(1-\sigma(Vh))}{\partial(1-\sigma(Vh))} \times \frac{\partial(1-\sigma(Vh))}{\partial Vh} \times \frac{\partial Vh}{\partial U}$

So, $\frac{\partial(A+B)}{\partial U} = (y-\sigma(Vh))\frac{\partial Vh}{\partial U}$

Since U is an $Sk \times F$ matrix and Vh is a scalar:

$$\frac{\partial Vh}{\partial U} = \begin{pmatrix} \frac{\partial Vh}{\partial u_{1,1}} & \cdots & \frac{\partial Vh}{\partial u_{1,F}} \\ \vdots & \cdots & \vdots \\ \frac{\partial Vh}{\partial u_{Sk,1}} & \cdots & \frac{\partial Vh}{\partial u_{Sk,F}} \end{pmatrix}$$

Where each $\frac{\partial Vh}{\partial u_{i,j}} = \frac{\partial(v_1 \cdot \max(p_1)+...+v_F \cdot \max(p_F))}{\partial u_{i,j}}$

since $u_{i,j}$ only appears in vector $\overrightarrow{u_j}$, which only appears in $\max(p_j)$

we only need to find $\frac{\partial v_j \cdot \max p_j}{\partial u_{i,j}}$

If we know the vector x_m s.t. $\tanh x_m^T u_j = \max(p_j)$ for each $j = 1, \cdots, F$

then: $\frac{\partial v_j \cdot \max p_j}{\partial u_{i,j}} = \frac{\partial v_j \tanh x_{m_j}^T u_j}{\partial x_{m_j}^T u_j} \times \frac{\partial x_{m_j}^T u_j}{u_{i,j}} = v_j(1-\tanh^2(x_{m_j}^T u_j))(x_{i,m_j})$

so:

$$\frac{\partial Vh}{\partial U} = \begin{pmatrix} v_1 x_{1,m_1}(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F x_{1,m_F}(1-\tanh^2(x_{m_F}^T u_F)) \\ \vdots & \cdots & \vdots \\ v_1 x_{Sk,m_1}(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F x_{Sk,m_F}(1-\tanh^2(x_{m_F}^T u_F)) \end{pmatrix}$$

$$\frac{\partial Vh}{\partial U} = \frac{\partial Vh}{\partial U} = \begin{pmatrix} v_1 \vec{w}_{m_1}(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F \vec{w}_{m_F}(1-\tanh^2(x_{m_F}^T u_F)) \\ \vdots & \cdots & \vdots \\ v_1 \vec{w}_{m_1+k-1}(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F \vec{w}_{m_F+k-1}(1-\tanh^2(x_{m_F}^T u_F)) \end{pmatrix}$$

$$\begin{pmatrix} v_1(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F(1-\tanh^2(x_{m_F}^T u_F)) \\ \vdots & \cdots & \vdots \\ v_1(1-\tanh^2(x_{m_1}^T u_1)) & \cdots & v_F(1-\tanh^2(x_{m_F}^T u_F)) \end{pmatrix}$$

For each filter $f = 1, \cdots, F$

4 Gradient Checking

V difference: 0.03543645, U difference: 0.00466077