Hao Zhang haomoodzhang@gmail.com

南京大學

- How to backprop on batch-normalization?

Sol: forward pass:

$$\mu_j \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)}, \ \forall i, j$$

$$v_j \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} (x_j^{(i)} - \mu_j)^2 = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)2} - \left(\frac{1}{m}\sum_{i=1}^{m} x_j^{(i)}\right)^2, \ \forall j$$

$$s_j^{(i)} = \frac{x_j^{(i)} - \mu_j}{\sqrt{v_j + \varepsilon}}, \ \forall i, j$$

$$a_j^{(i)} = \gamma_j s_j^{(i)} + \beta_j, \ \forall i, j$$

backward pass:

$$\frac{\partial \mu_j}{\partial x_j^{(i)}} = \frac{1}{m}$$

$$\frac{\partial v_j}{\partial x_j^{(i)}} = \frac{1}{m} 2 x_j^{(i)} - 2\left(\frac{1}{m}\sum_{i=1}^{m} x_j^{(i)}\right)\frac{1}{m} = \frac{2}{m}(x_j^{(i)} - \mu_j)$$

$$\cancel{\frac{\partial a_j^{(i)}}{\partial \mu_j} = \frac{1}{\sqrt{v_j+\varepsilon}}} \quad \frac{\partial J}{\partial \mu_j} = \sum_{i=1}^{m} \frac{\partial J}{\partial s_j^{(i)}} \frac{\partial s_j^{(i)}}{\partial \mu_j} = -\sum_{i=1}^{m} \frac{1}{\sqrt{v_j+\varepsilon}} \frac{\partial J}{\partial s_j^{(i)}}$$

$$\frac{\partial J}{\partial v_j} = \sum_{i=1}^{m} \frac{\partial J}{\partial s_j^{(i)}} \frac{\partial s_j^{(i)}}{\partial v_j} = -\sum_{i=1}^{m} \frac{1}{2} \frac{(x_j^{(i)} - \mu_j)}{(v_j+\varepsilon)^{\frac{3}{2}}} \frac{\partial J}{\partial s_j^{(i)}}$$

$$\therefore \frac{\partial J}{\partial x_j^{(i)}} = \frac{\partial J}{\partial s_j^{(i)}} \frac{\partial s_j^{(i)}}{\partial x_j^{(i)}} + \frac{\partial J}{\partial \mu_j} \frac{\partial \mu_j}{\partial x_j^{(i)}} + \frac{\partial J}{\partial v_j} \frac{\partial v_j}{\partial x_j^{(i)}}$$

$$= \frac{\partial J}{\partial s_j^{(i)}} \left(\frac{1}{\sqrt{v_j+\varepsilon}}\right) + \frac{1}{m}\frac{\partial J}{\partial \mu_j} + \frac{2}{m}(x_j^{(i)} - \mu_j)\frac{\partial J}{\partial v_j}$$