Gradients and Computational Graph

Aditya Chopra

October 2021

1 Computational Graph

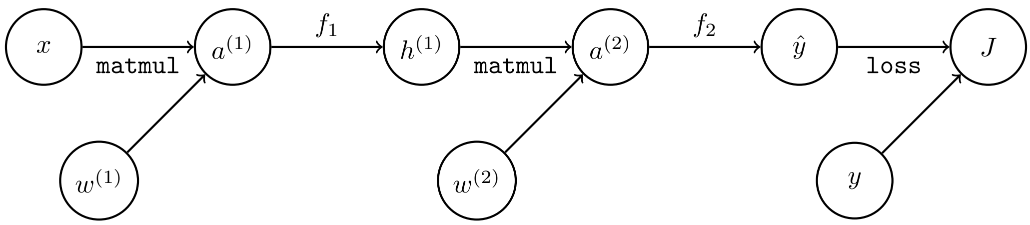

Figure 1: Reference Computational Graph for 2 Layer DNN

2 Forward Propagation

2.1 Shapes

$$
\begin{aligned}
x &\in \mathbb{R}^D \\
w^{(1)} &\in \mathbb{R}^{D\times M} \\
w^{(2)} &\in \mathbb{R}^{M\times K} \\
y &\in \mathbb{R}^K
\end{aligned}
$$

2.2 Equations

$$a^{(1)} = w^{(1)T} \cdot x \qquad \left[a^{(1)} \in \mathbb{R}^M\right] \tag{1}$$

$$h^{(1)} = f_1(a^{(1)}) \qquad \left[h^{(1)} \in \mathbb{R}^M\right] \tag{2}$$

$$a^{(2)} = w^{(2)T} \cdot h^{(1)} \qquad \left[a^{(2)} \in \mathbb{R}^K\right] \tag{3}$$

$$\hat{y} = f_2(a^{(2)}) \qquad \left[\hat{y} \in \mathbb{R}^K\right] \tag{4}$$

$$J = \mathcal{L}(y, \hat{y}) \tag{5}$$

3 Backward Propagation

3.1 Gradient of a scalar with respect to a vector

$$x \in \mathbb{R}^m \quad y \in \mathbb{R}^n \quad z \in \mathbb{R} \tag{6}$$

Abstraction 1:

$$(\nabla_x z)_i = \frac{\partial z}{\partial x_i} \tag{7}$$

Chain Rule:

$$\nabla_x z = \left(\frac{\partial y}{\partial x}\right)^T \nabla_y z \tag{8}$$

Where, $\frac{\partial y}{\partial x}$ is the Jacobian Matrix and

$$\frac{\partial y}{\partial x} \in \mathbb{R}^{n \times m} \tag{9}$$

3.2 Gradient of a scalar with respect to a tensor

X and Y are tensors in multiple dimensions.
Abstraction 1:

$$(\nabla_X z)_i = \frac{\partial z}{\partial X_i} \tag{10}$$

Chain Rule:

$$\nabla_X z = \sum_j (\nabla_X Y_j) \frac{\partial z}{\partial Y_j} \tag{11}$$

3.3 Equations

$$\nabla_{\hat{y}} J = \nabla_{\hat{y}} \mathcal{L}(\hat{y}, y) \qquad \left[\nabla_{\hat{y}} J \in \mathbb{R}^K\right] \tag{12}$$

$$\nabla_{a^{(2)}} J = \left(\frac{\partial \hat{y}}{\partial a^{(2)}}\right)^T \nabla_{\hat{y}} J \qquad \left[\frac{\partial \hat{y}}{\partial a^{(2)}} \in \mathbb{R}^{K \times K}, \quad \nabla_{a^{(2)}} J \in \mathbb{R}^K\right] \tag{13}$$

$$\nabla_{w^{(2)}} J = \sum_j \left(\nabla w^{(2)} Y_j\right) \frac{\partial J}{\partial Y_j} \qquad [] \tag{14}$$

$$\tag{15}$$