Ex 1: Gradients of Expected Loss

$$F(x, w), \quad G(x) \qquad x, w \in \mathbb{R}$$

$$y = F(G(x, w))$$

$$E = F(G(x^{(1)}, w)) + F(G(x^{(2)}, w))$$

Let $g^{(i)} = G(x^{(i)}, w)$ & $f^{(i)} = F(g^{(i)})$

Write an expression for $\frac{\partial E}{\partial w}$.

$$\frac{\partial E}{\partial w} = \frac{1}{D} \sum_{i=1}^{D} \frac{\partial f^{(i)}}{\partial w}$$

$$= \frac{1}{D} \sum_{i=1}^{D} \frac{\partial f^{(i)}}{\partial g^{(i)}} \frac{\partial g^{(i)}}{\partial w}$$

$$= \frac{1}{D} \sum_{i} \frac{\partial F(G(x^{(i)}, w))}{\partial G(x^{(i)}, w)} \frac{\partial G(x^{(i)}, w)}{\partial w} \qquad = \frac{1}{D} \sum_{i=1}^{D} F'\left(G(x^{(i)}, w)\right) \frac{\partial G(x^{(i)}, w)}{\partial w}$$

NOT

$$= F'\left(\sum_{i} \frac{\partial G(x^{(i)}, w)}{\partial w} \right)$$

Notation

$h^{(l)} = \sigma(z^{(l)})$

$h_i^{(l)}$ is i^{th} node in layer l

$h_{i,j}^{(l)}$ is j^{th} sample for $h_i^{(l)}$

$h_{i,\cdot}^{(l)}$ is all samples for $h_i^{(l)}$

$h_{\cdot,j}^{(l)}$ is act. of layer l for sample j

$= h^{(l)}$