Backpropagation in Full Connected Layers

f is activation

$$E = f(\hat{y} - y) = \frac{1}{2}\sum(\hat{y} - y)^2$$

$$z = xw + b$$

$$\frac{\partial E}{\partial w} = (\hat{y} - y)\frac{\partial(\hat{y} - y)}{\partial w} = (\hat{y} - y)\frac{\partial y}{\partial w}$$

$$= (\hat{y} - y)\frac{\partial f(z)}{\partial w} = (\hat{y} - y)\, f'(z)\frac{\partial z}{\partial w}$$

$$\frac{\partial E}{\partial w} = (\hat{y} - y) \cdot f'(z) \cdot x$$

---

with chain Rule

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial w} = \frac{\partial E}{\partial y} \cdot \frac{\partial f(z)}{\partial w}$$

$$(\hat{y} - y) \cdot f'(z) \cdot x$$