Common derivative term $\frac{\partial L}{\partial w_j} = (y^{(i)} - g^{(i)}) x^{(i)}_j$ [from 1 to m]

Update rules

1. SGD $dw = \alpha \frac{\partial L}{\partial w}$

2. SGD with momentum $dw = \alpha \frac{\partial L}{\partial w} + \mu dw$

3. Adagrad $sum = sum + \left(\frac{\partial L}{\partial w}\right)^2$

$$dw = \frac{\alpha\left(\frac{\partial L}{\partial w}\right)}{\sqrt{sum}}$$

4. RMS prop $sum = (sum)\,0.9 + \left(\frac{\partial L}{\partial w}\right)^2 (0.1)$

$$dw = \frac{\alpha\left(\frac{\partial L}{\partial w}\right)}{\sqrt{sum}}$$