SGD weight update:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \gamma \nabla_{\mathbf{w}^{(t)}} f(\mathbf{w}^{(t)})$$
$$f(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} f_i(\mathbf{w})$$
$$f_i = \ell(\mathbf{w}^T \mathbf{x}_i, y_i)$$

Define $s = \mathbf{w}^T \mathbf{x}$, so $\frac{\partial s}{\partial \mathbf{w}} = \mathbf{x}$.

$$\nabla f = \frac{1}{n}\sum_{i=1}^{n} \nabla f_i$$
$$\nabla_{\mathbf{w}} f_i(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \ell(\mathbf{w}^T \mathbf{x}_i, y_i) = \frac{\partial}{\partial s} \ell(s, y_i)\mathbf{x_i}$$

Loss functions are hinge, squared and logistic.

$$\ell_{hinge} = \max\{0, 1 - (\mathbf{w}^T, \mathbf{x}_i) y_i\}$$
$$\ell_{squared} = \frac{1}{2}||y_i - \mathbf{w}^T \mathbf{x}_i||_2^2$$
$$\ell_{logistic} = -\log(1 + \exp(-(\mathbf{w}^T \mathbf{x}_i) y_i))$$

Hence, the partial derivatives:

$$\frac{\partial}{\partial s}\ell_{hinge} = \begin{cases} -y_i, & \text{if } (\mathbf{w}^T \mathbf{x}_i) y_i < 1 \\ 0, & \text{if } (\mathbf{w}^T \mathbf{x}_i) y_i \geq 1 \end{cases}$$
$$\frac{\partial}{\partial s}\ell_{squared} = (\mathbf{w}^T \mathbf{x}_i - y_i)$$
$$\frac{\partial}{\partial s}\ell_{logistic} = \frac{y_i}{1 + \exp(-(\mathbf{w}^T \mathbf{x}_i) y_i)}$$

For ℓ_1 and ℓ_2 regularizers

$$\Omega_1(\mathbf{w}) = \lambda ||\mathbf{w}||_1 \implies \frac{\partial}{\partial \mathbf{w}} \Omega_1(\mathbf{w}) = \lambda\, sign(\mathbf{w})$$
$$\Omega_2(\mathbf{w}) = \frac{\lambda}{2}||\mathbf{w}||_2^2 \implies \frac{\partial}{\partial \mathbf{w}} \Omega_2(\mathbf{w}) = \lambda \mathbf{w}$$