

$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, b) = 2(\hat{\mathbf{y}} - \mathbf{y})\mathbf{X}^T$$