

Backpropagation

The model computes predictions using

$$\hat{y} = wx + b$$

The loss function (MSE) is

$$L = \frac{1}{n} \sum (y - \hat{y})^2$$

To reduce the loss, we compute gradients using the chain rule.

Gradient w.r.t. weights:

$$\frac{\partial L}{\partial w} = -\frac{2}{n} \sum x(y - \hat{y})$$

Gradient w.r.t. ~~weight~~ bias:

$$\frac{\partial L}{\partial b} = -\frac{2}{n} \sum (y - \hat{y})$$

weights are updated using gradient descent:

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}$$

$$b \leftarrow b - \eta \frac{\partial L}{\partial b}$$

The gradients flow from the loss backward to the parameters, which is why the process is called backpropagation.