Gradient for linear regression

Pawel Wocjan

January 16, 2020

Abstract

We consider the gradient for linear regression for the simplest case.

1 Gradient

Let $w \in \mathbb{R}$ and $b \in \mathbb{R}$ be the weight and bias for linear regression. Given $x \in \mathbb{R}$, the predicted value is

$$\hat{y} = wx + b. \tag{1}$$

Assume that the correct value for x is $y \in \mathbb{R}$. Then the squared error loss is given by

$$\mathcal{L} = \frac{1}{2}(\hat{y} - y)^2. \tag{2}$$

The gradient of the loss function is

$$\nabla \mathcal{L} = \begin{pmatrix} \dfrac{\partial \mathcal{L}}{\partial w} \\ \\ \dfrac{\partial \mathcal{L}}{\partial b} \end{pmatrix} \tag{3}$$

Using the chain rule, we obtain the weight component of the gradient

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = (\hat{y} - y) \cdot x. \tag{4}$$

The expression for the bias component of the gradient is even simpler

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} = (\hat{y} - y). \tag{5}$$