Gradient Decent Based Polynomial Regression

Loic Guegan

February 14, 2021

First, choose a polynomial function $h_w(x)$ according to the data complexity. In our case, we have:

$$h_w(x) = w_1 + w_2x + w_3x^2 \tag{1}$$

Then, we should define a cost function. A common approach is to use the **Mean Square Error** cost function:

$$J(w) = \frac{1}{2n}\sum_{i=0}^{n}(h_w(x^{(i)}) - y^{(i)})^2 \tag{2}$$

With n the number of observations, $x^{(i)}$ the value of the independant variable associated with the observation $y^{(i)}$. Note that in Equation 2 we average by $2n$ and not n. This is because it simplify the partial derivatives expression as we will see below. This is a pure cosmetic approach which do not impact the gradient decent (see here for more informations). The next step is to $min_wJ(w)$ for each weight w_i (performing the gradient decent, see here). Thus we compute each partial derivatives:

$$\begin{aligned}\frac{\partial J(w)}{\partial w_1} &= \frac{\partial J(w)}{\partial h_w(x)}\frac{\partial h_w(x)}{\partial w_1}\\ &= \frac{1}{n}\sum_{i=0}^{n}(h_w(x^{(i)}) - y^{(i)})\end{aligned} \tag{3}$$

similarly:

$$\frac{\partial J(w)}{\partial w_2} = \frac{1}{n}\sum_{i=0}^{n}x(h_w(x^{(i)}) - y^{(i)}) \tag{4}$$

$$\frac{\partial J(w)}{\partial w_3} = \frac{1}{n}\sum_{i=0}^{n}x^2(h_w(x^{(i)}) - y^{(i)}) \tag{5}$$