Batch Gradient Descent Algorithm for Linear Regression

- Assume 1-D dataset
- Cost Function:

$$J(\mathbf{w}) = MSE = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

where

$$\hat{y}_i = f\left(x^{(i)}\right) = w_0 + w_1 x^{(i)}$$

Training

repeat until *convergence* {

// Update equation

$$w_j = w_j - \eta \, \frac{\partial J(\mathbf{w})}{\partial w_j}$$

$j = 0, \dots, n$ where n is the number of features

$$temp0 = w_0 - \eta \, \frac{\partial J(w_0, w_1)}{\partial w_0} = w_0 - \eta . \frac{1}{m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y_i\right)$$

$$temp1 = w_1 - \eta \, \frac{\partial J(w_0, w_1)}{\partial w_1} = w_1 - \eta . \frac{1}{m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y_i\right) . x^{(i)}$$

$$w_0 = temp0$$

$$w_1 = temp1$$

m is the number of training examples

η is the learning rate

// Compute MSE

$$MSE = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 = \frac{1}{2m}\sum_{i=1}^{m}\left(w_0 + w_1 x^{(i)} - y_i\right)^2$$

}

Prediction

$$\hat{y} = f(x) = w_0 + w_1 x$$

Where $[w_0 \quad w_1]$ is the optimal parameter vector and x is the test instance