Batch Gradient Descent Algorithm for Logistic Regression

- Assume 4-D dataset

$$\mathbf{w}^T = [w_0 \quad w_1 \quad w_2 \quad w_3 \quad w_4]$$
$$\mathbf{x}^T = [x_0 \quad x_1 \quad x_2 \quad x_3 \quad x_4] \text{ where } x_0 = 1$$

$$\sigma(\mathbf{w}^T\mathbf{x}) = 1/\big(1 + \exp\big(-(w_0x_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4)\big)\big)$$

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

$\frac{\partial J(\mathbf{w})}{\partial w_j} = \frac{1}{m}\sum_{i=1}^{m}\left(\left(\sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right) - y_i\right).\, x_j^{(i)}\right)$

$$temp0 = w_0 - \eta \, \frac{\partial J(\mathbf{w})}{\partial w_0} = w_0 - \eta . \frac{1}{m}\sum_{i=1}^{m}\left(\left(\sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right) - y_i\right).\, x_0^{(i)}\right) \text{ where } x_0^{(i)} = 1$$

$$temp1 = w_1 - \eta \, \frac{\partial J(\mathbf{w})}{\partial w_1} = w_1 - \eta . \frac{1}{m}\sum_{i=1}^{m}\left(\left(\sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right) - y_i\right).\, x_1^{(i)}\right)$$

$$\cdots$$

$$temp4 = w_4 - \eta \, \frac{\partial J(\mathbf{w})}{\partial w_4} = w_4 - \eta . \frac{1}{m}\sum_{i=1}^{m}\left(\left(\sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right) - y_i\right).\, x_4^{(i)}\right)$$

$$w_0 = temp0, w_1 = temp1, \dots, w_4 = temp4$$

m is the number of training examples

η is the learning rate

// Compute the cost function

$$J = -\sum_{i=1}^{m}\left(\left(y_i \log\left(\sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right)\right)\right) + \left((1 - y_i)\log\left(1 - \sigma\left(\mathbf{w}^T\mathbf{x}^{(i)}\right)\right)\right)\right)$$

}

Prediction

$$\text{if } \sigma(\mathbf{w}^T\mathbf{x}) \geq 0.5 \rightarrow \hat{y} = 1 \text{ else } \hat{y} = 0$$

where $\mathbf{w}$ is the optimal weight vector obtained after training and $\mathbf{x}$ is the test instance.