CS3390 Assignment 1
Problem 4

Gautam Singh (CS21BTECH11018)
Jaswanth Beere (BM21BTECH11007)

Contents

This document discusses an algorithm to find a solution to a binary classification problem that uses the sigmoid function in logistic regression.

1 Newton-Raphson Update Equation

Define

$$p_i \triangleq \Pr\left(y_i = 1 | \mathbf{x}_i, \mathbf{w}\right) = \frac{1}{1 + \exp\left(-\mathbf{w}^\top \mathbf{x}_i\right)} \tag{1}$$

The likelihood function for logistic regression is given by

$$L\left(\mathbf{w}\right) = \prod_{i:y_i=1} p_i \prod_{j:y_j=0} \left(1 - p_j\right) \tag{2}$$

Hence, the log likelihood is given by

$$l\left(\mathbf{w}\right) = \sum_{i=1}^{N} y_i \log p_i + \left(1 - y_i\right) \log\left(1 - p_i\right) \tag{3}$$

and using (1), the error function is simply the negative of (3), given by

$$E\left(\mathbf{w}\right) = -\sum_{i=1}^{N} y_i \log p_i + \left(1 - y_i\right) \log\left(1 - p_i\right) \tag{4}$$

$$= -\sum_{i=1}^{N} y_i \left(\mathbf{w}^\top \mathbf{x}_i\right) - \log\left(1 + \exp\left(\mathbf{w}^\top \mathbf{x}_i\right)\right) \tag{5}$$

Using (1), the first derivatives of (5) are

$$\frac{\partial E}{\partial w_r} = -\sum_{i=1}^{N} \left(y_i - \frac{1}{1 + \exp\left(-\mathbf{w}^\top \mathbf{x}_i\right)}\right) x_{ir} \tag{6}$$

$$= -\sum_{i=1}^{N} \left(y_i - p_i\right) x_{ir}. \tag{7}$$

The second derivatives are therefore

$$\frac{\partial^2 E}{\partial w_r \partial w_s} = \sum_{i=1}^{N} x_{ir} x_{is} \frac{\exp\left(-\mathbf{w}^\top \mathbf{x}\right)}{\left(1 + \exp\left(-\mathbf{w}^\top \mathbf{x}_i\right)\right)^2} \tag{8}$$

$$= \sum_{i=1}^{N} x_{ir} x_{is} p_i \left(1 - p_i\right). \tag{9}$$

Hence, from (7) the *gradient* of (5) is

$$\nabla E\left(\mathbf{w}\right) = -\sum_{i=1}^{N} \left(y_i - p_i\right) \mathbf{x}_i = -\mathbf{X}^\top \left(\mathbf{y} - \mathbf{p}\right) \tag{10}$$

and from (9), the Hessian becomes

$$\mathbf{H}_E\left(\mathbf{w}\right) = \mathbf{X}^\top \mathbf{W} \mathbf{X}, \tag{11}$$

where we define

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \end{pmatrix} \tag{12}$$

$$\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \dots & y_N \end{pmatrix}^\top \tag{13}$$

$$\mathbf{p} = \begin{pmatrix} p_1 & p_2 & \dots & p_N \end{pmatrix}^\top \tag{14}$$

$$\mathbf{W} = \operatorname{diag}\begin{pmatrix} p_1\left(1 - p_1\right) & \dots & p_N\left(1 - p_N\right) \end{pmatrix}. \tag{15}$$

Thus, the Hessian update equation becomes

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mathbf{H}_E\left(\mathbf{w}_n\right)^{-1} \nabla E\left(\mathbf{w}_n\right) \tag{16}$$

$$= \mathbf{w}_n + \left(\mathbf{X}^\top \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^\top \left(\mathbf{y} - \mathbf{p}\right). \tag{17}$$

Therefore, the optimal solution is

$$\hat{\mathbf{w}}_{ML} = \left(\mathbf{X}^\top \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^\top \left(\mathbf{y} - \mathbf{p}\right). \tag{18}$$

Note that the optimal solution is quite similar to the one for reweighted least squares regression in Problem 3, where $\mathbf{y} - \mathbf{p}$ is the "corrected" outputs with the probabilities of being the correction.