Foundations of Machine Learning: Assignment 2 Attachment

Newton's Method

Consider the problem of finding the root of a equation $f(x) = 0$. Each iteration of Newton's method is

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}.$$

Repeat this procedure until the algorithm converges and then a approximated solution can be acquired.

Iterative Reweighted Least Squares

The logistic regression optimization problem is to maximize the conditional likelihood

$$\mathcal{L}(\boldsymbol{w}) = \sum_i \left[y_i \boldsymbol{w}^T \boldsymbol{x}_i - \log(1 + exp(\boldsymbol{w}^T \boldsymbol{x}_i)) \right]$$

It can be proved that this function is concave. We need to find $\boldsymbol{w}^*$ such that

$$\nabla \mathcal{L}(\boldsymbol{w}^*) = 0.$$

Similar to Newton's method, we can perform the following iteration

$$\boldsymbol{w}_{t+1} \leftarrow \boldsymbol{w}_t - H^{-1} \nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})|_{\boldsymbol{w}_t},$$

where H is Hessian matrix:

$$H = \nabla^2_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})|_{\boldsymbol{w}_t}.$$

The gradient is

$$\nabla_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})|_{\boldsymbol{w}_t} = \sum_i (y_i - \mu_i) \boldsymbol{x}_i = X(\boldsymbol{y} - \boldsymbol{\mu})$$

$$\mu_i = 1/(1 + exp(-\boldsymbol{w}_t^T \boldsymbol{x}_i))$$

The Hessian matrix is:

$$H = \nabla^2_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w})|_{\boldsymbol{w}_t} = -\sum_i \mu_i (1 - \mu_i) \boldsymbol{x}_i \boldsymbol{x}_i^T = -XRX^T,$$

where $R_{ii} = \mu_i(1 - \mu_i)$. Thus the iteration can be written as:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t + (XRX^T)^{-1} X(\boldsymbol{y} - \boldsymbol{\mu})$$