Iteratively Reweighted Least Squares (IRLS)

IRLS is a class of estimation methods for regression-based problems. For generalized linear models, it is almost equivalent to Newton's method. This is the default estimation method for `glm.fit` in R. In general, there is no general recipe for an IRLS method. Potentially, if a computational method consists of steps involving 1) weighted least square estimation and 2) iteratively updating/re-evaluating the weights, it may be considered as an IRLS method.

1 IRLS for ℓ_1 regression and ℓ_p regression, $p < 1$, in compressed sensing

The ℓ_1 regression is to find $\boldsymbol{\beta}$ which minimizes $\arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1 = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} |y_i - \mathbf{x}_i\boldsymbol{\beta}|,$.

Re-write the loss $\sum_{i=1}^n |y_i - \mathbf{x}_i\boldsymbol{\beta}| = \sum_{i=1}^n \frac{1}{|y_i-\mathbf{x}_i\boldsymbol{\beta}|}(y_i - \mathbf{x}_i\boldsymbol{\beta})^2 = \sum_{i=1}^n w_i(\boldsymbol{\beta})(y_i - \mathbf{x}_i\boldsymbol{\beta})^2$, where $w_i(\boldsymbol{\beta}) = \frac{1}{|y_i-\mathbf{x}_i\boldsymbol{\beta}|}$.

1. Initialize $\boldsymbol{\beta}^{(0)}$ (which may be a least-square estimate)
2. Update $\boldsymbol{\beta}^{(t+1)}$ as the weighted least-square estimate based on the weighted least-square criteria $\boldsymbol{\beta}^{(t+1)} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n w_i(\boldsymbol{\beta}^{(t)})(y_i - \mathbf{x}_i\boldsymbol{\beta})^2$. The solution is $\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{y})$, where $\mathbf{W}^{(t)}$ is a diagonal matrix with (i,i)-th entry is $w_i(\boldsymbol{\beta}^{(t)})$.
3. The weights here are based on the value of $\boldsymbol{\beta}$ from the previous iteration. Thus, it is called 'Iteratively Reweighted', and the weighted least square is applied afterward to get the solution given these weights.

In compressed sensing, the problem is to find the parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ which minimize the ℓ_p norm for the linear regression problem, $\arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_p = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} |y_i - \mathbf{x}_i\boldsymbol{\beta}|^p$, where $p < 1$.

2 IRLS for Generalized linear model fitting

We learn this by example. The strategy is essentially the same. We need to simplify the second derivative and write the NR update as a least-square solution.

Let $(y_i, \mathbf{x}_i)_{1\le i\le n}$ be the set of observations where $y_i \in \{0,1\}$. The logistic regression model is,

$$
\begin{aligned}
y_i \sim& \text{Bernoulli}(p_i)\\
p_i =& \frac{1}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})}
\end{aligned}
$$

Log-likelihood $= \sum_{i=1}^n [y_i \text{log} p_i + (1-y_i)\text{log}(1-p_i)]$.

Thus, $f(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^n [y_i \text{log} p_i + (1-y_i)\text{log}(1-p_i)]$. This is the loss function or objective function.

$f'(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\{y_i(1-p_i) - (1-y_i)p_i\} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i(y_i - p_i)$.

$f''(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n -\mathbf{x}_i\frac{\partial p_i}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{\{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})\}^2} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T p_i(1-p_i) = \frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{X}$, where $\mathbf{X}$ is a $n\times p$ matrix whose i-th row is $\mathbf{x}_i$ and $\mathbf{W}$ is a diagonal matrix with (i,i)-th entry is $p_i(1-p_i)$.

Now rewriting the first derivative as, $f'(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i p_i(1-p_i)\frac{y_i-p_i}{p_i(1-p_i)} = \frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{z}$, where $\mathbf{z} = \{\frac{y_i-p_i}{p_i(1-p_i)}\}_{1\le i\le n} = \mathbf{W}^{-1}(\mathbf{y}-\mathbf{p})$

Note that $\mathbf{W}$ and $\mathbf{z}$ both depend on $\boldsymbol{\beta}$. In case of NR updating, $\mathbf{W}$ and $\mathbf{z}$ would depend on $\boldsymbol{\beta}^{(t)}$ while getting the update $\boldsymbol{\beta}^{(t+1)}$. So we write $\mathbf{W}^{(t)}$ and $\mathbf{z}^{(t)}$

Finally, the NR update is $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{z}^{(t)} = (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})\boldsymbol{\beta}^{(t)} - (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{z}^{(t)} = (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(t)}(\mathbf{X}\boldsymbol{\beta}^{(t)} - \mathbf{z}^{(t)})$

Thus, we get our IRLS update as,

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(t)}(\mathbf{X}\boldsymbol{\beta}^{(t)} - \mathbf{z}^{(t)}) = (\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(t)}\mathbf{u}^{(t)}$$

1. Initialize your $\boldsymbol{\beta}^{(0)}$
2. Get updated dummy responses $\mathbf{u}^{(t)} = \mathbf{X}\boldsymbol{\beta}^{(t)} - \mathbf{z}^{(t)}$ and get the $\mathbf{W}^{(t)}$.
3. Apply the weighted least square solution.
4. Repeat until convergence.

We will re-do this for Poisson GLM in the homework.