Iteratively Reweighted Least Squares

On HW6, we developed a majorize-minimize algorithm to solve

$$\min_{x \in \mathbb{R}^D} \|Ax-b\|_1 = \sum_{i=1}^{N} |a_i^T x - b_i|.$$

We will now take a direct approach to solving this problem using Iteratively Reweighted Least Squares (IRLS).

(Notation change)

Let $e = Ax-b$ and consider minimizing

$$\|Ax-b\|_p = \|e\|_p = \left(\sum_{i=1}^{N} |e_i|^p\right)^{1/p}$$

where $p \in (0,2)$. Note that

$$|e_i|^p = |e_i|^{p-2}|e_i|^2$$
$$= w_i^2 |e_i|^2$$

where we substitute $w_i^2 = |e_i|^{p-2}$. Taking the sum over i and noting that we can instead minimize $\|e\|_p^p$, our problem becomes

$$\min_{x \in \mathbb{R}^D} \sum_{i=1}^{N} w_i^2 |e_i|^2 = \min_x \sum_{i=1}^{N} w_i^2 (a_i^T x - b_i)^2$$

where $w_i = |e_i|^{\frac{p-2}{2}}$. This is a weighted least-squares problem, which we know how to solve! This leads to the following, more general algorithm for robust regression. The superscript denotes the iteration

Alg 1: IRLS for ℓ_p minimization

$x^0 = A^+ b$

while not converged

$e^k = Ax^{k-1} - b$

$w_i^k = |e_i|^{\frac{p-2}{2}}$ for $i = 1, \dots, N$ (element-wise)

$$W = \begin{bmatrix} w_1^k & & \\ & \ddots & \\ & & w_N^k \end{bmatrix} \in \mathbb{R}^{N \times N}$$

$$x^k = \min_x \| W(Ax - b) \|_2^2$$

IRLS for Sparse Regression

It is also possible to use IRLS to solve the following equality-constrained sparse regression

$$\min_{x \in \mathbb{R}^D} \|x\|_p \quad \text{ST } Ax = b. \qquad (1)$$

Since sparse regression is often applied when $A \in \mathbb{R}^{N \times D}$ is wide $(N \ll D)$, the equality constraint can typically be satisfied in practice.

To solve (1) using IRLS, we replace $\|x\|_1$ with $x^T W x$ for some diagonal $W \in \mathbb{R}^{D \times D}$, since

$$\|x\|_p^p = \sum_{i=1}^{D} |x_i|^p = \sum_{i=1}^{D} |x_i|^2 \underbrace{|x_i|^{p-2}}_{w_i} = x^T W x .$$

In practice, we add a bit of regularization by setting

$$\boxed{w_i = \left(x_i^2 + \varepsilon\right)^{\frac{p-2}{2}}, \quad i = 1, \dots, D}$$

for some small $\varepsilon > 0$. This leads to the equality constrained problem

$$\min_{x \in \mathbb{R}^D} \; x^T W x \qquad \text{ST} \quad Ax = b . \qquad (2)$$

The solution to (2) can be found using the method of Lagrange multipliers, which you may (should?) have seen in your undergraduate calculus course. First, write the Lagrangian as

$$\mathcal{L}(x, \lambda) = x^T W x + \lambda^T (Ax - b) .$$

From optimization theory, we know that the optimal x and λ must satisfy

$$\nabla_\lambda \mathcal{L}(x, \lambda) = Ax - b = 0 \in \mathbb{R}^N$$

$$\nabla_x \mathcal{L}(x, \lambda) = 2Wx + A^T \lambda = 0 \in \mathbb{R}^D$$

The second step is to solve for x using the second equation above.

$$2Wx = -A^T\lambda \iff x = -\frac{1}{2}W^{-1}A^T\lambda$$

(Why is W invertible?) Now plug this value of x into the first equation above to get

$$Ax = b \iff A\left(-\frac{1}{2}W^{-1}A^T\lambda\right) = b$$

$$\iff \left(-\frac{1}{2}AW^{-1}A^T\right)\lambda = b$$

$$\iff \lambda = \left(-\frac{1}{2}AW^{-1}A^T\right)^{-1}b.$$

Finally, substitute this value to solve for the optimal x

$$x = -\frac{1}{2}W^{-1}A^T\lambda$$

$$= -\frac{1}{2}W^{-1}A^T\left(-\frac{1}{2}AW^{-1}A^T\right)^{-1}b$$

$$= \boxed{W^{-1}A^T\left(AW^{-1}A^T\right)^{-1}b}$$

The resulting algorithm is given below.

Alg 2: IRLS for sparse regression

$$x^0 = \text{zeros}(D, 1)$$

while not converged

$$w_i^k = \left((x_i^{k-1})^2 + \varepsilon\right)^{\frac{p-2}{2}} \qquad i = 1, \dots, D$$

$$W = \begin{bmatrix} w_1^k & & \\ & \ddots & \\ & & w_D^k \end{bmatrix}$$

$$x^k = W^{-1} A^T (A W^{-1} A^T)^{-1} b$$

Convergence Criteria

How do we know when to stop the above algorithms?
For Alg 1, a reasonable choice is when

$$\|A x^k - b\|_2 < \varepsilon \qquad (3)$$

for some small $\varepsilon \sim 10^{-6}$. For Alg 2, we are enforcing the $Ax^k = b$ constraint, so (3) is a bad choice.
A better idea is to stop when

$$\|x^k - x^{k-1}\| < \varepsilon$$

which is an indication that the algorithm has converged.