Gradient Descent for Locally Weighted Linear Regression

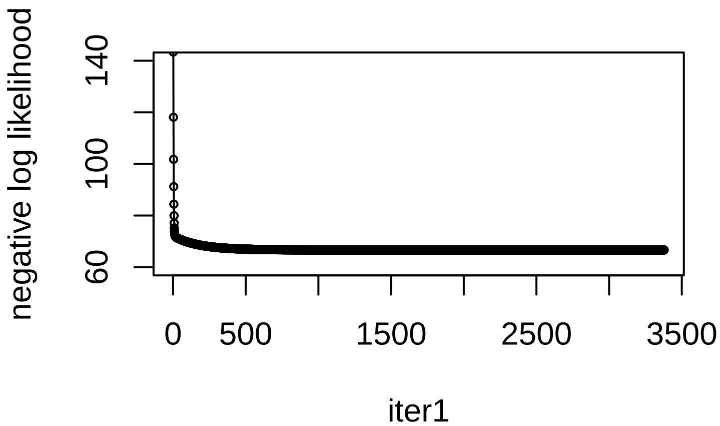
Wenchen Guo

Gradient Descent for Unweighted Linear Regression.

Use data rx.dat and ry.dat, which contain the predictors x_i and response y_i respectively. Implement gradient descent for (unweighted) linear regression.

Set the threshold to be 10^{-10} and step_size = 0.005.

The log-likelihood function versus iterations plot:



Intercept	theta
0.3277	0.1753

Gradient descent runs 3380 iterations and the log likelihood function value converge to 66.6729.
The coefficients are $(0.3277, 0.1753)$.

Locally Weighted Linear Regression.

Consider a linear regression problem in which we want to weight different training examples differently. Therefore, we want to minimize

$$\begin{aligned} \mathrm{J}(\theta) &= \frac{1}{2}\sum_{i=1}^{n} w_i \left(\theta^T x_i - y_i\right)^2 \\ &= (X\theta - y)^T W (X\theta - y) \end{aligned}$$

where

$$\theta = (\theta_1 \ldots \theta_p)^T, \qquad y = (y_1 \ldots y_n)^T$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11} & \ldots & x_{p1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \ldots & x_{pn} \end{bmatrix}, \qquad W = \frac{1}{2}\begin{bmatrix} w_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & w_n \end{bmatrix}$$

Suppose we have samples $(x_i, y_i), i = 1, ..., n$ of n independent examples, but in which the y_i's were observed with different variances, and

$$p(y_i|x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma_i^2}\right)$$

i.e. y_i has mean $\theta^T x_i$ and variance σ_i^2 (where σ_i^2 are fixed, known, constants).

Find the loglikelihood function:

$$l(p(y_i)) = -log\left(\sqrt{2\pi\sigma_i^2}\right) - \frac{(y_i - \theta^T x_i)^2}{2\sigma_i^2}$$

$$l(\prod_{i=1}^{n} p(y_i)) = -\sum_{i=1}^{n} log\left(\sqrt{2\pi\sigma_i^2}\right) - \sum_{i=1}^{n} \frac{(y_i - \theta^T x_i)^2}{2\sigma_i^2}$$

Since variance σ_i^2 are fixed, known, constants, the first term is constant.
Thus, finding the maximum likelihood estimater of θ is equalvelant to find the *argmin* of the second term, which is to find the *argmin* of

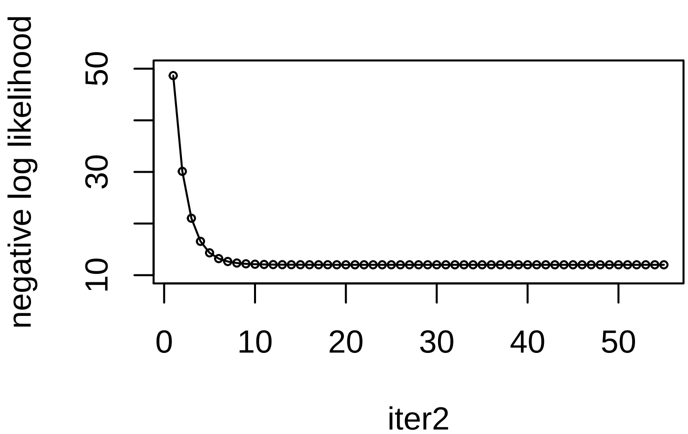
$$\mathrm{J}(\theta) = \frac{1}{2}\sum_{i=1}^{n} w_i \left(\theta^T x_i - y_i\right)^2$$

taking $w_i = W_{ii} = \frac{1}{\sigma_i^2}$

Gradient Descent for Locally Weighted Linear Regression.

Set weight as $w_i = \exp\left(-x_i^2/20\right)$, and set the threshold to be 10^{-10} and step_size = 0.005.

The log-likelihood function versus iterations plot:



Intercept	theta
0.3943	0.4157

Gradient descent runs 56 iterations and the log likelihood function value converge to 12.0112.
The coefficients are $(0.3943, 0.4157)$.

Comparison.

Fit result plot: