HW2

Yilun Zhang

October 9, 2015

Exercise 1

Clearly, if the convergence rate is Q-quadratic, then it's Q-superlinear. Since if

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||^2} \leq \gamma$$

then

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{||x_{k+1} - x^*||}{||x_k - x^*||^2} \times ||x_k - x^*|| \leq \gamma ||x_k - x^*|| \to 0$$

a

$x^* = 0$

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{\frac{1}{k+1}}{\frac{1}{k}} = \frac{k}{k+1} < 1$$

So the rate is Q-linear.

b

$x^* = 1$

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||^2} = \frac{0.5^{2^{k+1}}}{(0.5^{2^k})^2} = 1$$

So the rate is Q-quadratic.

c

$x^* = 1$
$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{(k+1)^{-(k+1)}}{k^{-k}} \leq \frac{(k+1)^k}{(k+1)^{k+1}} = \frac{1}{k+1} \to 0$$
The rate is Q-superlinear.

d

$x^* = 0$
$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{k!}{(k+1)!} = \frac{1}{k+1} = \to 0$$
The rate is Q-superlinear.

e

$x^* = 0$. If k is even
$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{(1/4)^{2k}}{k(1/4)^{2k}} = 1/k \to 0$$
If k is odd,
$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} = \frac{k(1/4)^{2^{k+1}}}{(1/4)^{2^{k-1}}} = k(1/4)^{2^{k+1}-2^{k-1}} \to 0$$
The rate is Q-superlinear.

Exercise 2

a

By chain rule:
$$\bigtriangledown h(y) = \bigtriangledown f(x) \cdot \frac{\partial (Sy)}{\partial y} = S' \bigtriangledown f(x)$$
$$\bigtriangledown^2 h(y) = \bigtriangledown S' \bigtriangledown^2 f(x) S$$

b

By Newton's method:

$$\begin{aligned} y_{k+1} &= y_k - (\bigtriangledown^2 h(y_k))^{-1} \bigtriangledown h(y) \\ &= y_k - S^{-1} \bigtriangledown^2 f(x_k)(S')^{-1} S' \bigtriangledown f(x_k) \\ &= y_k - S^{-1} \bigtriangledown^2 f(x_k) \bigtriangledown f(x_k) \end{aligned}$$

Let $y_k = S^{-1} x_k$ and mutiply S to both side:

$$x_{k+1} = x_k - \bigtriangledown^2 f(x_k) \bigtriangledown f(x_k)$$

This is the Newton's formula of x_k

Exercise 3

a

The goal is to minimize the negeative log likelihood.

$$l(x, y, \theta) = -\sum_{i=1}^{n} [y_i log p_i + (1 - y_i) log(1 - p_i)]$$

where

$$p_i = \frac{e^{\theta^T x}}{1 + e^{\theta^T x}}$$

The Jacobian is

$$J_k = \frac{\partial l(x, y, \theta)}{\theta_k} = \sum_{i=1}^{n} (y_i x_{ik} - p_i x_{ik})$$

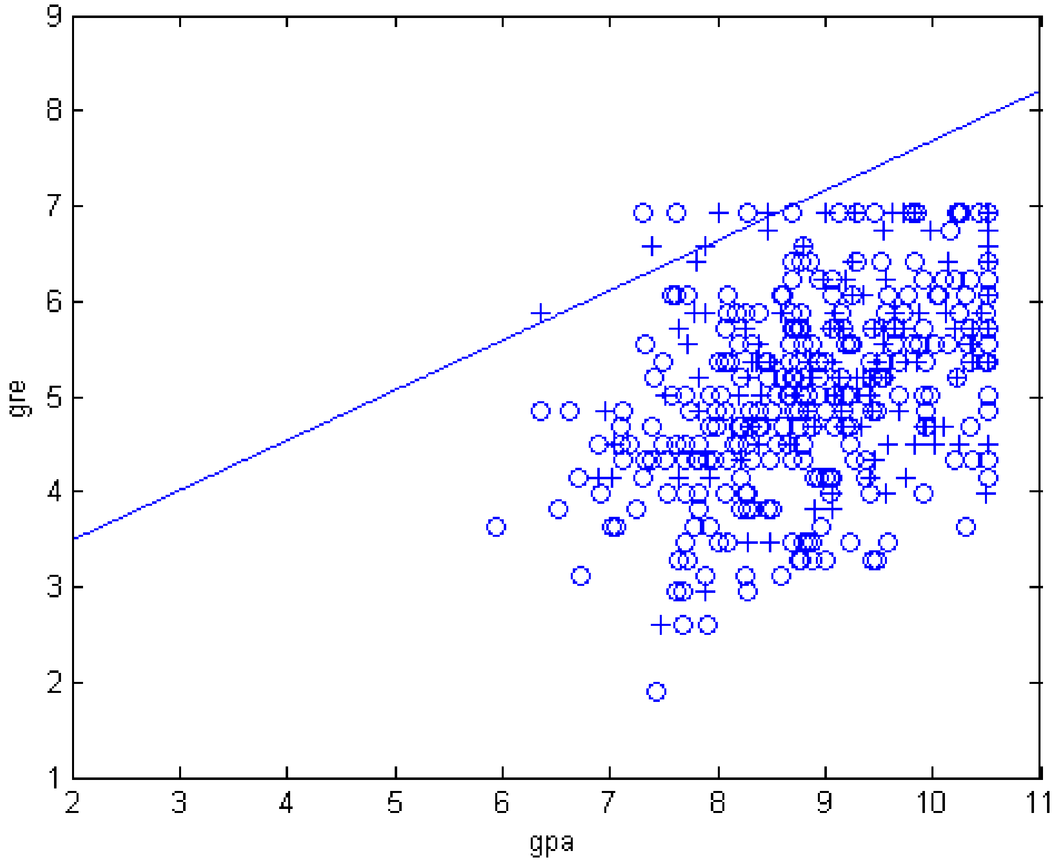
Therefore, the gradient descent method is

$$\theta_{k+1} = \theta_k - \alpha J$$

b

Standardize the data before fitting the model. See codes and plot.

c&d



gre
gpa