Solution to Homework 2

March 21, 2016

1. By the strong convexity of $J(\theta)$, we have that

$$J(\theta) \geq J(\theta^*) + \nabla J(\theta^*)^T(\theta - \theta^*) + \frac{\delta}{2}\|\theta - \theta^*\|^2$$

Since θ^* is the minimizer, we have $\nabla J(\theta^*) = 0$ and $J(\theta) - J(\theta^*) \geq 0$, then

$$\|\theta - \theta^*\|^2 \leq \frac{2}{\delta}(J(\theta) - J(\theta^*)) \tag{1}$$

On the other hand, $J(\theta)$ is also convex, then

$$J(\theta^*) \geq J(\theta) + \nabla J(\theta)^T(\theta^* - \theta)$$

i.e.

$$0 \leq J(\theta) - J(\theta^*) \leq \nabla J(\theta)^T(\theta - \theta^*)$$

so we can obtain

$$J(\theta) - J(\theta^*) \leq \|\nabla J(\theta)\|\|\theta - \theta^*\| \leq \sqrt{\frac{\delta\epsilon}{2}}\|\theta - \theta^*\|$$

i.e.

$$(J(\theta) - J(\theta^*))^2 \leq \frac{\delta\epsilon}{2}\|\theta - \theta^*\|^2 \tag{2}$$

Combining (1) and (2) together, we conclude that

$$J(\theta) - J(\theta^*) \leq \epsilon.$$

2.

(i)

Algorithm 1 Gradient descent algorithm

Input: X, y, θ_0, α, tol

Output: θ, F, g

$k \leftarrow 0$

while $\frac{\|\nabla f(\theta^k)\|}{\max\{1, \|\nabla f(\theta^0)\|\}} > \text{tol}$ **do**

$\theta^{k+1} \leftarrow \theta^k - \alpha \nabla f(\theta^k)$

$F(k+1) = f(\theta^{k+1})$

$g(k+1) = \|\nabla f(\theta^{k+1})\|_2$

$k \leftarrow k+1$

end while

return θ, F, g

(ii) The stepsize can be chosen as $0 \leq \alpha \leq \frac{1}{L} \approx 0.033$, where $L = \lambda_{max}(X^T X)$. The solution is $\theta = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$.

(iii) Linear convergence.

(iv) You can use the stepsize as $0 \leq \alpha \leq \frac{1}{L} \approx 7.99 \times 10^{-6}$.

(v) Sublinear convergence