Gosuddin Siddiqi - HW1

Student No: 1627383

Problem 1) (1)

The given function is,

$$f(x) = \frac{1}{2}\,||\mathbf{F}x - \mathbf{r}||^2$$

Computing the ∇f for the above function, we get,
∇f = $\frac{1}{2}\,(2.||\mathbf{F}x - \mathbf{r}||).\,||\mathbf{F}||$
= $(||\mathbf{F}x - \mathbf{r}||).\,||\mathbf{F}||$
=$(||\mathbf{F}^T.\,\mathbf{F}x - \mathbf{F}^T.\,\mathbf{r}||)\ \dots (1)$
= $\mathbf{F}^T(\mathbf{F}x - \mathbf{r})\ \dots (2)$
Therefore, ∇f = $\mathbf{F}^T(\mathbf{F}x - \mathbf{r})$

Further, computing the $\nabla^2 f$ from the equation (1),
= $\mathbf{F}^T.\,\mathbf{F}\ \dots (3)$

The Lipschitz continuity condition is given by,

$||f(x) - f(y)||_2 \leq L|\mathbf{F}(x) - \mathbf{F}y|$

Using the gradient from equation (2),

= $||\mathbf{F}^T(\mathbf{F}x - \mathbf{r}) - \mathbf{F}^T(\mathbf{F}y - \mathbf{r})||$
= $||\mathbf{F}^T.\,\mathbf{F}x - \mathbf{F}^T.\,\mathbf{r}\ \ - \mathbf{F}^T.\,\mathbf{F}y + \mathbf{F}^T.\,\mathbf{r}||$
= $||\mathbf{F}^T.\,\mathbf{F}||_2\,||\mathbf{x} - \mathbf{y}||_2$

$\because$ ∇f is C^1 and satisfies the above condition, we say, the function $\nabla f(x)$ is Lipschitz continuous with Lipschitz constant = $||\mathbf{F}^T.\,\mathbf{F}||$

Problem 1) (2)

Similarly, for,

$$f(x) = \frac{1}{2}\,||\mathbf{F}x - \mathbf{r}||^2 + \frac{\lambda}{2}\,||x||^2$$

From equation (2) we get,
∇f = $\mathbf{F}^T(\mathbf{F}x - \mathbf{r}) + \lambda x\ \dots (4)$

From equation (4), we get,
$\nabla^2 f = \mathbf{F}^T.\mathbf{F} + \lambda \dots (5)$

Using the definition of Lipschitz continuity stated above and equation (4), we get,
$= ||\mathbf{F}^T(\mathbf{F}x - \mathbf{r}) + \lambda x - \mathbf{F}^T(\mathbf{F}y - \mathbf{r}) - \lambda y||$
$= (||\mathbf{F}^T.\mathbf{F}||_2 + \lambda)(||\mathbf{x} - \mathbf{y}||_2)$

Again,
$\nabla f(x)$ is C^1 and satisfies the Lipschitz continuous condition, it is Lipschitz continuous with Lipschitz constant = $||\mathbf{F}^T.\mathbf{F}||_2 + \lambda$

$$\min_x f(x) = \frac{1}{2}\,\|\mathbf{F}x - \mathbf{r}\|^2 + \lambda\|x\|_1$$

Problem 2) (1)

Let,
$\frac{1}{2}\,\|\mathbf{F}x - \mathbf{r}\|^2$ be part (1)

and $\lambda\|x\|_1$ be part (2)

Since, $\lambda||x||_1$ is not C^1 even though part (1) of $f(x)$ is β - smooth, the function $f(x)$ is not entirely β-smooth.

Problem 2) (2)

The give equation is,

$$\min_x f(x) := \frac{1}{2\eta}\,\|x - y\|^2 + \lambda\|x\|_1$$

Expanding the $\|x - y\|^2$ term,

$= \min_x \frac{1}{2\eta}\,[(\|x\|^2 - 2.(x.y) - \|y\|^2)] + \lambda\|x\|_1 \dots (1)$

Since y is independent of x, we can exclude it from the minimizer and treat each value of x as x_i because x_i is independent of $x_i \pm 1$.

Our equation (1) becomes,

$= \min_x \frac{1}{2\eta}\,[\Sigma_i x_i^2 - 2.(x_i.y_i)] + \lambda\Sigma_i|x_i|$

$= \min_x \frac{1}{\eta}\,[\Sigma_i\,\frac{x_i^2}{2} - (x_i.y_i)] + \lambda\Sigma_i|x_i|$

$= \min_x \frac{1}{\eta}\,[\Sigma_i\,\frac{x_i^2}{2} - (x_i.y_i) + \lambda\eta|x_i|] \dots (2)$

Now all the terms from equation can be treated as individual minimization problem as they are individual data points of the dataset and can be solved for closed form.

Rewriting the equation (2) for derivative,

$= \frac{x_i^2}{2} - (x_i . y_i) + \lambda\eta x_i \cdots (3)$

When $y_i > 0, x_i \geq 0$, equation (3),

$= \frac{x_i^2}{2} - (x_i . y_i) + \lambda\eta x_i \cdots (4)$

When $y_i < 0, x_i \leq 0$, equation (3),

$= \frac{x_i^2}{2} - (x_i . y_i) - \lambda\eta x_i \cdots (5)$

Taking the derivative of $(4), (5)$

$= 2.x_i - y_{\pm}\lambda\eta$

Equating to 0 and solving for x,

$0 = 2.x_i - y_{\pm}\lambda\eta$

$x_i = y_i \mp \lambda\eta \ldots. (6)$

Thus, based on the sign of y_i we can come up with an expression,

$x_i = sgn(y_i)(|y_i| - \lambda\eta)$

Hence,

$$\min_x f(x) := \frac{1}{2\eta} \|x - y\|^2 + \lambda\|x\|_1$$

can be minimized with following minimizer,

$$x_i = sgn(y_i)(|y_i| - \lambda\eta)$$

Problem 3)(1)

We let,

$$p_k(a) = min_x \frac{1}{2} (x - a)^2 + k|a|$$

From our solution in 2 (2), we get,

$p_k(a) = \frac{1}{2} [sgn(a)(|a| - k) - a]^2 + k(||a| - k|)$

Now, when $|a| > k$,

$p_k(a) = \frac{1}{2}k^2 + k|a - sgn(a)k|$

$= \frac{1}{2}k^2 + k||a| - k|$

$= k|a| - \frac{1}{2}k^2 \dots (1)$

Further, $|a| < k$,

$p_k(a) = \frac{a^2}{2} \dots (2)$

By combining equation (1) and (2),

$p_k(a)$ confines with the definition of $\rho_k(a)$

Problem 3)(2)

Calculating the derivative of ρ_k, we get,

When $|a| > k$,
$k.sgn(a)$

and when $|a| < k$,
a

Thus, we conclude that $\rho_k(a)$ is C^1 continuous and β-smooth.

Using the Lipschitz continuity condition mentioned in problem 1,

when $|a| \leq k$, values, $x, y \leq k$,

we get, $||F(x) - F(y)|| = |y - x|$

Similarly we can prove this for other cases as well. Hence, ρ_k is β smooth, with β as 1.

Problem 3)(3)

These are two functions coupled together and both are C^1 and hence β smooth.

Let,
$f(x) = \rho_k(Fx - r)$

Computing derivative, we get,

$\nabla f(x) = F^T \nabla \rho_k(Fx - r)$

Using Lipschitz continuity, we get,

$\|F(x) - F(y)\|_2 < \|F^T.F\|\|(x-y)\|_2$

Hence, we get $\beta = F^T.F$

Problem 4)(1)

The given equation is,

$$\min_x f(x) := \sum_{i=1}^{m} \left\{ \log\left(1 + \exp\left\langle f^i, x \right\rangle\right) - s^i \left\langle f^i, x \right\rangle \right\} + \frac{\lambda}{2}\|x\|^2$$

Let,
$u = \left\langle f^i, x \right\rangle$

$f(x)$ becomes,

$\sum_{i=1}^{m} \log(1 + \exp(u)) + \frac{\lambda}{2}\|x\|^2 \ldots (1)$

We know that, derivative of $\log(1 + \exp(u)) = \frac{\exp(u)}{1+\exp(u)}$

Substituting in equation (1) , we get,
$\nabla f(x) = \sum_{i=1}^{m} u^{`}.exp(u)/(1 + \exp(u)) + \frac{\lambda}{2}\|x\|^2$

$exp(u)/(1 + \exp(u))$ can be expressed as a sigmoid function,

$\frac{1}{1+exp(-u)}$

Introducing the sigmoid function, we get,
$\nabla f(x) = F^T(\frac{1}{1+\exp(-u)} - s) + \lambda x$

Problem 4)(2)

We can see that $\nabla f(x)$ is C^1 continuous, hence it is β-smooth

Using the Lipschitz continuity,

we can prove that, similar to linear regression problem in 1,
we get, $\beta = F^t.F + \lambda$