ANLY561 Homework7

Hongyang Zheng

Question 1

$$\nabla(g \circ f)(x) = \begin{pmatrix} \frac{\partial(g \circ f)}{\partial f} \frac{\partial f}{\partial x1}(x) \\ \frac{\partial(g \circ f)}{\partial f} \frac{\partial f}{\partial x2}(x) \\ . \\ . \\ . \\ \frac{\partial(g \circ f)}{\partial f} \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

$$= g'(f(x)) \begin{pmatrix} \frac{\partial f}{\partial x1}(x) \\ \frac{\partial f}{\partial x2}(x) \\ . \\ . \\ . \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

$$= g'(f(x))\nabla f(x)$$

Question 2

a)

$$\nabla f_1(x)^T = \begin{pmatrix} \frac{\partial f1}{\partial x_1}(x) & \frac{\partial f1}{\partial x_2}(x) & \dots & \frac{\partial f1}{\partial x_n}(x) \end{pmatrix}$$

$$\nabla f_2(x)^T = \begin{pmatrix} \frac{\partial f2}{\partial x_1}(x) & \frac{\partial f2}{\partial x_2}(x) & \dots & \frac{\partial f2}{\partial x_n}(x) \end{pmatrix}$$

$$.$$
$$.$$
$$.$$

$$\nabla f_m(x)^T = \begin{pmatrix} \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

Therefore, we have:

$$
Df(x) = \begin{pmatrix} \nabla f_1(x)^T \\ \nabla f_2(x)^T \\ . \\ . \\ . \\ \nabla f_m(x)^T \end{pmatrix}
$$

b)

$$
\nabla(g \circ f)(x) = \begin{pmatrix} \frac{\partial(g \circ f_1)}{\partial f_1} \frac{\partial f_1}{\partial x_1}(x) + \frac{\partial(g \circ f_2)}{\partial f_2} \frac{\partial f_2}{\partial x_1}(x) + \ldots + \frac{\partial(g \circ f_m)}{\partial f_m} \frac{\partial f_m}{\partial x_1}(x) \\ \frac{\partial(g \circ f_1)}{\partial f_1} \frac{\partial f_1}{\partial x_2}(x) + \frac{\partial(g \circ f_2)}{\partial f_2} \frac{\partial f_2}{\partial x_2}(x) + \ldots + \frac{\partial(g \circ f_m)}{\partial f_m} \frac{\partial f_m}{\partial x_2}(x) \\ . \\ . \\ . \\ \frac{\partial(g \circ f_1)}{\partial f_1} \frac{\partial f_1}{\partial x_n}(x) + \frac{\partial(g \circ f_2)}{\partial f_2} \frac{\partial f_2}{\partial x_n}(x) + \ldots + \frac{\partial(g \circ f_m)}{\partial f_m} \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_1}(x) & \ldots & \frac{\partial f_m}{\partial x_1}(x) \\ \frac{\partial f_1}{\partial x_2}(x) & \frac{\partial f_2}{\partial x_2}(x) & \ldots & \frac{\partial f_m}{\partial x_2}(x) \\ . & . & . & \\ . & . & & . \\ . & . & & & . \\ \frac{\partial f_1}{\partial x_n}(x) & \frac{\partial f_2}{\partial x_n}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \begin{pmatrix} \frac{\partial(g \circ f_1)}{\partial f_1} \\ \frac{\partial(g \circ f_2)}{\partial f_2} \\ . \\ . \\ . \\ \frac{\partial(g \circ f_m)}{\partial f_m} \end{pmatrix}
$$

$$
= Df(x)^T \nabla g(f(x))
$$

(c)

Given $(g \circ f) \in C^1(R^n, R^m)$

$$
D(g \circ f)(\mathbf{x}) = \begin{pmatrix} \frac{\partial(g \circ f)_1}{\partial x_1}(\mathbf{x}) & \frac{\partial(g \circ f)_1}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial(g \circ f)_1}{\partial x_n}(\mathbf{x}) \\ . \\ . \\ . \\ \frac{\partial(g \circ f)_m}{\partial x_1}(\mathbf{x}) & \frac{\partial(g \circ f)_m}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial(g \circ f)_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}
$$

$g \in C^1(R^k, R^m), f \in C^1(R^n, R^k)$

$$D(g \circ f)(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial f_1}\frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial g_1}{\partial f_1}\frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_1}{\partial f_1}\frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ . & & & \\ . & & & \\ . & & & \\ \frac{\partial g_m}{\partial f_k}\frac{\partial f_k}{\partial x_1}(\mathbf{x}) & \frac{\partial g_m}{\partial f_k}\frac{\partial f_k}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial g_m}{\partial f_k}\frac{\partial f_k}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Since

$$Df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ . & & & \\ . & & & \\ . & & & \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

only values on diagonal line exist, like $\frac{\partial f_1}{\partial x_1}, \frac{\partial f_2}{\partial x_2} \dots \frac{\partial f_m}{\partial x_n}$ exist, others are regarded as derivative on constants and will be 0,

Therefore, we have

$$Dg(f(\mathbf{x})) = \begin{pmatrix} \frac{\partial g_1}{\partial f_1}(\mathbf{x}) & \dots & \frac{\partial g_1}{\partial f_k}(\mathbf{x}) \\ \frac{\partial g_2}{\partial f_1}(\mathbf{x}) & \dots & \frac{\partial g_2}{\partial f_k}(\mathbf{x}) \\ . & & \\ . & & \\ . & & \\ \frac{\partial g_m}{\partial f_1}(\mathbf{x}) & \dots & \frac{\partial g_m}{\partial f_k}(\mathbf{x}) \end{pmatrix}$$

and

$$Df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & 0 & \dots & 0 \\ 0 & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \dots & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

Therefore, $D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x})$.

3.

(a)

$$l(\widetilde{\beta}) = -\sum_{i=1}^{N} loglogit(y_i(\widetilde{\mathbf{x}}^{(i)})^T\widetilde{\beta}) = -\{loglogit(y_1(\widetilde{\mathbf{x}}^{(1)})^T\widetilde{\beta} + \ldots .\, loglogit(y_n(\widetilde{\mathbf{x}}^{(n)})^T\widetilde{\beta}\}$$

$$= (1, 1, 1, \ldots, 1)\begin{pmatrix} -loglogit(y_1(\widetilde{\mathbf{x}}^{(1)})^T\widetilde{\beta}) \\ -loglogit(y_2(\widetilde{\mathbf{x}}^{(2)})^T\widetilde{\beta}) \\ . \\ . \\ . \\ -loglogit(y_n(\widetilde{\mathbf{x}}^{(n)})^T\widetilde{\beta}) \end{pmatrix}$$

$$= s(\begin{pmatrix} -loglogit(y_1(\widetilde{\mathbf{x}}^{(1)})^T\widetilde{\beta}) \\ -loglogit(y_2(\widetilde{\mathbf{x}}^{(2)})^T\widetilde{\beta}) \\ . \\ . \\ . \\ -loglogit(y_n(\widetilde{\mathbf{x}}^{(n)})^T\widetilde{\beta}) \end{pmatrix})$$

$$= s(\begin{pmatrix} -loglogit(y_1 q_1) \\ -loglogit(y_2 q_2) \\ . \\ . \\ . \\ -loglogit(y_n q_n) \end{pmatrix}), q_i = (\widetilde{\mathbf{x}}^{(i)})^T\widetilde{\beta}$$

So

$$l(\widetilde{\beta}) = s(g((\widetilde{\mathbf{x}}^{(i)})^T\widetilde{\beta}),$$

since $f(\widetilde{\beta}) = \mathbf{X}\widetilde{\beta}$ for

$$\mathbf{X} = \begin{pmatrix} (\widetilde{\mathbf{x}}^{(1)})^T\widetilde{\beta} \\ (\widetilde{\mathbf{x}}^{(2)})^T\widetilde{\beta} \\ . \\ . \\ . \\ (\widetilde{\mathbf{x}}^{(n)})^T\widetilde{\beta} \end{pmatrix}$$

Therefore, $(\widetilde{\mathbf{x}}^{(i)})^T\widetilde{\beta} = f(\widetilde{\beta})$ and $l(\widetilde{\beta}) = s(g(f(\widetilde{\beta})))$

(b)

We have

$$\nabla l(\widetilde{\beta}) = \begin{pmatrix} \partial_1(\mathbf{1}^T{}_N g(f(\widetilde{\beta}))) \\ \partial_2(\mathbf{1}^T{}_N g(f(\widetilde{\beta}))) \\ . \\ . \\ . \\ \partial_n(\mathbf{1}^T{}_N g(f(\widetilde{\beta}))) \end{pmatrix}$$

Using matrix product rule,

$$\partial_n(\mathbf{1}^T{}_N g(f(\widetilde{\beta}))) = \partial_n(\mathbf{1}^T{}_N)(g(f(\widetilde{\beta}))) + (\mathbf{1}^T{}_N)\partial_n(g(f(\widetilde{\beta})))$$

$\mathbf{1}^T{}_N$ is the vector with all entries all equal to 1, so

$$\partial_n(\mathbf{1}^T{}_N g(f(\widetilde{\beta})) = \partial_n(g(f(\widetilde{\beta})))$$

$$\partial_n g(f(\widetilde{\beta})) = \partial_n(-loglogit(y_n X_n \widetilde{\beta}_n)) = -X_n y_n\, logit(-y_n X_n \widetilde{\beta}_n) = X_n h(f(\widetilde{\beta}_n))$$

Therefore,

$$\nabla l(\widetilde{\beta}) = \begin{pmatrix} X_1 h(f(\widetilde{\beta}_1)) \\ X_2 h(f(\widetilde{\beta}_2)) \\ . \\ . \\ . \\ X_n h(f(\widetilde{\beta}_n)) \end{pmatrix} = X^T h(f(\widetilde{\beta}))$$

(c)

$$\nabla^2 l(\widetilde{\beta}) = \begin{pmatrix} \partial_{1,1} X_1 h(f(\widetilde{\beta}_1)) & \partial_{1,2} X_1 h(f(\widetilde{\beta}_1)) & \dots & \partial_{1,n} X_1 h(f(\widetilde{\beta}_1)) \\ \partial_{1,2} X_2 h(f(\widetilde{\beta}_2)) & \partial_{2,2} X_2 h(f(\widetilde{\beta}_2)) & \dots & \partial_{2,n} - X_2 h(f(\widetilde{\beta}_2)) \\ . & & & \\ . & & & \\ . & & & \\ \partial_{1,n} X_n h(f(\widetilde{\beta}_n)) & \partial_{2,n} X_n h(f(\widetilde{\beta}_n)) & \dots & \partial_{n,n} - X_n h(f(\widetilde{\beta}_n)) \end{pmatrix}$$

The values on the diagonal line exist, others will be 0, pick one value:

$$\partial_n - X_n y_n logit(-y_n X_n \widetilde{\beta}_n) = (-y_n X_n) - X_n y_n logit(-(-y_n X_n \widetilde{\beta}_n)) logit(-y_n X_n \widetilde{\beta}_n) =$$

$$X_n^2 y_n^2 logit(y_n X_n \widetilde{\beta}_n) logit(-y_n X_n \widetilde{\beta}_n)$$

Given y_n equals either 1 and -1,

$$\partial_n - X_n y_n logit(-y_n X_n \widetilde{\beta}_n) = X_n logit(X_n \widetilde{\beta}_n) logit(-X_n \widetilde{\beta}_n) X_n$$

$$\nabla^2 l(\widetilde{\beta}) = \left(\begin{matrix} X_1 logit(X_1 \widetilde{\beta}_1) logit(-X_1 \widetilde{\beta}_1) X_1 & \dots & 0 \\ 0 & X_2 logit(X_2 \widetilde{\beta}_2) logit(-X_2 \widetilde{\beta}_2) X_2 & \dots \\ . & & \\ . & & \\ . & & \\ 0 & \dots & X_n logit(X_n \widetilde{\beta}_n) logit(-X_n \widetilde{\beta}_n) X \end{matrix} \right.$$

$$(X_1, X_2, \dots, X_n) \begin{pmatrix} logit(X_1 \widetilde{\beta}_1) logit(-X_1 \widetilde{\beta}_1) & 0 & \dots & 0 \\ 0 & logit(X_2 \widetilde{\beta}_2) logit(-X_2 \widetilde{\beta}_2) & \dots & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & 0 & \dots & logit(X_n \widetilde{\beta}_n) logit(-X_n \widetilde{\beta}_n) \end{pmatrix}$$

$$\begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{pmatrix}$$

Therefore, $\nabla^2 l(\widetilde{\beta}) = X^T diag(G(f(\widetilde{\beta})))X$

(d)

$\nabla^2 l(\widetilde{\beta})^T = \nabla^2 l(\widetilde{\beta})$, therefore $\nabla^2 l(\widetilde{\beta}) \in M_{n,n}$ is symmetric.

$G(f(\widetilde{\beta})) = logit(X\widetilde{\beta})logit(-X\widetilde{\beta}) > 0$ for all $\widetilde{\beta}$,$diag(G(f(\widetilde{\beta}))) > 0$ for all $\widetilde{\beta}$

$\nabla^2 l(\widetilde{\beta}) \geq 0$, and $z^T \nabla^2 l(\widetilde{\beta}) z \geq 0$ for all $z \in R^n$.

Therefore, $\nabla^2 l(\widetilde{\beta})$ is always positive semidefinite.

(e)

$\nabla^2 l(\widetilde{\beta}) = X^T diag(G(f(\widetilde{\beta})))X$, by theorem "Eigenvalue Characterization of Positive Definiteness", $\nabla^2 l(\widetilde{\beta})$ is positive definite if and only if it has strictly positive eigenvalues.

By spectral theorem, if $\nabla^2 l(\widetilde{\beta})$ is in the form of $U^T diag(v) U$ where U is orthogonal, then v list the eigenvalues of $\nabla^2 l(\widetilde{\beta})$.

$diag(G(f(\widetilde{\beta})))$ has strictly positive values in each entry, so if X is orthogonal, the conditions are satisfied.

Therefore, $\nabla^2 l(\widetilde{\beta})$ is positive definite when X is orthogonal.

(f)

Part (d) implies that $l(\widetilde{\beta})$ is convex.

Part (e) satisfied, then $l(\widetilde{\beta})$ is strictly convex.

When $\widetilde{\beta}^*$ satisfies $\nabla l(\widetilde{\beta}) = 0$ then it is the solution of $min_{\widetilde{\beta} \in R^n} l(\widetilde{\beta})$.