ANLY 561 HW

Name:Yuqi Wang NetID:yw545

Problem 1

Part (a)

$$\ell(ilde{eta}) = -\sum_{k=1}^N \log \operatorname{logit}(y_i \cdot (ilde{x}^{(i)})^T ilde{eta}) \ f(ilde{eta}) = X ilde{eta} ext{ for } X = egin{pmatrix} (ilde{x}^{(1)})^T \ (ilde{x}^{(2)})^T \ dots \ (ilde{x}^{(N)})^T \end{pmatrix} = egin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,n} \ 1 & x_{2,1} & \cdots & x_{2,n} \ dots & dots & \ddots & dots \ 1 & x_{N,1} & \cdots & x_{N,n} \end{pmatrix}$$

$$g(\mathbf{q}) = egin{pmatrix} -\log \operatorname{logit}(y_1q_1) \ -\log \operatorname{logit}(y_2q_2) \ dots \ -\log \operatorname{logit}(y_Nq_N) \end{pmatrix} \ s(\mathbf{v}) = \mathbf{1}_N^T \mathbf{v}$$

So

$$f(ilde{eta}) = egin{pmatrix} (ilde{x}^{(1)})^T ilde{eta} \ (ilde{x}^{(2)})^T ilde{eta} \ dots \ (ilde{x}^{(N)})^T ilde{eta} \end{pmatrix}$$

Then

$$g(f(ilde{eta})) = egin{pmatrix} - ext{log logit}(y_1 \cdot (ilde{x}^{(1)})^T ilde{eta}) \ - ext{log logit}(y_2 \cdot (ilde{x}^{(2)})^T ilde{eta}) \ dots \ - ext{log logit}(y_N \cdot (ilde{x}^{(N)})^T ilde{eta}) \end{pmatrix}$$

Finally, we have

$$s(g(f(ilde{eta}))) = \mathbf{1}_N^T g(f(ilde{eta})) = (\ 1 \ \ 1 \ \ \cdots \ \ 1) \left(egin{array}{c} -\mathrm{log} \ \mathrm{logit}(y_1 \cdot (ilde{x}^{(1)})^T eta) \ -\mathrm{log} \ \mathrm{logit}(y_2 \cdot (ilde{x}^{(2)})^T ilde{eta}) \end{array}
ight) \ dots \ -\mathrm{log} \ \mathrm{logit}(y_N \cdot (ilde{x}^{(N)})^T ilde{eta}) \end{array}
ight)$$

Thus

$$egin{aligned} s(g(f(ilde{eta}))) &= -\mathrm{log} \ \mathrm{logit}(y_1 \cdot (ilde{x}^{(1)})^T ilde{eta}) - \mathrm{log} \ \mathrm{logit}(y_2 \cdot (ilde{x}^{(2)})^T ilde{eta}) - \cdots - \mathrm{log} \ \mathrm{logit}(y_N \cdot (ilde{x}^{(N)})^T ilde{eta}) \ &= -\sum_{k=1}^N \mathrm{log} \ \mathrm{logit}(y_i \cdot (ilde{x}^{(i)})^T ilde{eta}) = \ell(ilde{eta}) \end{aligned}$$

Part (b)

Firstly, we can get the following properties by some simple calculations,

$$\mathrm{logit}(x) = rac{1}{1 + e^{-x}}, \ (\mathrm{log}\, \mathrm{logit}(x))' = \mathrm{logit}(-x)$$

And we define

$$h(\mathbf{q}) = egin{pmatrix} -y_1 \mathrm{logit}(-y_1q_1) \ -y_2 \mathrm{logit}(-y_2q_2) \ dots \ -y_N \mathrm{logit}(-y_Nq_N) \end{pmatrix}$$

From part (a), we have $\ell(ilde{eta}) = s(g(f(ilde{eta})))$, so

$$abla \ell(ilde{eta}) =
abla (s \circ g \circ f)(ilde{eta}) =
abla (s \circ (g \circ f))(ilde{eta})$$

Then we use multivariate chain rule,

$$abla (s\circ (g\circ f))(ilde{eta}) = D(g\circ f)(ilde{eta})^T
abla s(g\circ f(ilde{eta}))$$

We use multivariate chain rule again to get $D(g\circ f)(\tilde{\beta})=Dg(f(\tilde{\beta}))Df(\tilde{\beta})$, so $D(g\circ f)(\tilde{\beta})^T\nabla s(g\circ f(\tilde{\beta}))=Df(\tilde{\beta})^TDg(f(\tilde{\beta}))^T\nabla s(g(f(\tilde{\beta})))$

Since $f(ilde{eta}) = X ilde{eta}$, so

$$Df(ilde{eta}) = X, \ Df(ilde{eta})^T = X^T.$$

Then since
$$g(\mathbf{q}) = \begin{pmatrix} -\log \log \mathrm{it}(y_1q_1) \\ -\log \log \mathrm{it}(y_2q_2) \\ \vdots \\ -\log \log \mathrm{it}(y_Nq_N) \end{pmatrix}$$
 and $(\log \log \mathrm{it}(x))' = \mathrm{logit}(-x)$, so
$$Dg(\mathbf{q}) = \begin{pmatrix} -\log \mathrm{it}(-y_1q_1)y_1 & 0 & \cdots & 0 \\ 0 & -\log \mathrm{it}(-y_2q_2)y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\log \mathrm{it}(-y_Nq_N)y_N \end{pmatrix}$$

Thus

$$Dg(f(ilde{eta})) = egin{pmatrix} -\mathrm{logit}(-y_1(ilde{x}^{(1)})^T ilde{eta})y_1 & 0 & \cdots & 0 \ 0 & -\mathrm{logit}(-y_2(ilde{x}^{(2)})^T ilde{eta})y_2 & \cdots & 0 \ dots & dots & \ddots & dots \ 0 & 0 & \cdots & -\mathrm{logit}(-y_N(ilde{x}^{(N)})^T ilde{eta})y_N \end{pmatrix} = Dg(f(ilde{eta}))^T$$

Since $s(\mathbf{v}) = \mathbf{1}_N^T \mathbf{v}$, we have

$$abla s(\mathbf{v}) = \mathbf{1}_N$$

Also,

$$abla s(g(f(ilde{eta}))) = \mathbf{1}_N$$

Finally,

$$\nabla \ell(\tilde{\beta}) = Df(\tilde{\beta})^T Dg(f(\tilde{\beta}))^T \nabla s(g(f(\tilde{\beta})))$$

$$= X^T \begin{pmatrix} -\operatorname{logit}(-y_1(\tilde{x}^{(1)})^T \tilde{\beta}) y_1 & 0 & \cdots & 0 \\ 0 & -\operatorname{logit}(-y_2(\tilde{x}^{(2)})^T \tilde{\beta}) y_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\operatorname{logit}(-y_N(\tilde{x}^{(N)})^T \tilde{\beta}) y_N \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= X^T \begin{pmatrix} -y_1 \operatorname{logit}(-y_1(\tilde{x}^{(1)})^T \tilde{\beta}) \\ -y_2 \operatorname{logit}(-y_2(\tilde{x}^{(2)})^T \tilde{\beta}) \\ \vdots \\ -y_N \operatorname{logit}(-y_N(\tilde{x}^{(N)})^T \tilde{\beta}) \end{pmatrix}$$

Since
$$h(\mathbf{q}) = \left(egin{array}{c} -y_1 \mathrm{logit}(-y_1q_1) \ -y_2 \mathrm{logit}(-y_2q_2) \ dots \ -y_N \mathrm{logit}(-y_Nq_N) \end{array}
ight)$$
 , so

$$h(f(ilde{eta})) = egin{pmatrix} -y_1 \mathrm{logit}(-y_1(ilde{x}^{(1)})^T ilde{eta}) \ -y_2 \mathrm{logit}(-y_2(ilde{x}^{(2)})^T ilde{eta}) \ dots \ -y_N \mathrm{logit}(-y_N(ilde{x}^{(N)})^T ilde{eta}) \end{pmatrix}$$

Therefore,

$$abla \ell(ilde{eta}) = X^T h(f(ilde{eta}))$$

Part (c)

Firstly, we can get the following properties by some simple calculations,

$$ext{logit}(x) = rac{1}{1+e^{-x}}, \ ext{logit}'(x) = ext{logit}(x) ext{logit}(-x)$$

And we define

$$G(\mathbf{q}) = egin{pmatrix} \log \mathrm{it}(q_1) \mathrm{logit}(-q_1) \ \log \mathrm{it}(q_2) \mathrm{logit}(-q_2) \ dots \ \log \mathrm{it}(q_N) \mathrm{logit}(-q_N) \end{pmatrix}$$

$$\mathrm{diag}(d) = egin{pmatrix} d_1 & 0 & \cdots & 0 \ 0 & d_2 & \cdots & 0 \ dots & dots & \ddots & dots \ 0 & 0 & \cdots & d_N \end{pmatrix}$$

Observe that

$$\begin{split} \nabla^2 \ell(\tilde{\beta}) &= \nabla \nabla \ell(\tilde{\beta}) = D \nabla \ell(\tilde{\beta}) = D(X^T h(f(\tilde{\beta}))) \\ &= X^T D(h \circ f)(\tilde{\beta}) = X^T D(h(f(\tilde{\beta})) D(f(\tilde{\beta})) \text{ (Using multivariate chain rule)} \\ &= X^T D(h(f(\tilde{\beta})) X \text{ (In part b, we have showed that } D(f(\tilde{\beta})) = X) \end{split}$$

$$\text{Since } h(\mathbf{q}) = \begin{pmatrix} -y_1 \text{logit}(-y_1 q_1) \\ -y_2 \text{logit}(-y_2 q_2) \\ \vdots \\ -y_N \text{logit}(-y_N q_N) \end{pmatrix}, \text{ so } \\ Dh(\mathbf{q}) = \begin{pmatrix} y_1^2 \text{logit}(y_1 q_1) \text{logit}(-y_1 q_1) & 0 & \cdots & 0 \\ 0 & y_2^2 \text{logit}(y_2 q_2) \text{logit}(-y_2 q_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_N^2 \text{logit}(y_N q_N) \text{logit}(-y_N q_N) \end{pmatrix}$$

and
$$h(f(ilde{eta})) = egin{pmatrix} -y_1 \mathrm{logit}(-y_1(ilde{x}^{(1)})^T ilde{eta}) \ -y_2 \mathrm{logit}(-y_2(ilde{x}^{(2)})^T ilde{eta}) \ dots \ -y_N \mathrm{logit}(-y_N(ilde{x}^{(N)})^T ilde{eta}) \end{pmatrix}$$
 , so

$$=\begin{pmatrix} y_1^2 \mathrm{logit}(y_1(\tilde{x}^{(1)})^T \tilde{\beta}) \mathrm{logit}(-y_1(\tilde{x}^{(1)})^T \tilde{\beta}) & 0 & \cdots & 0 \\ 0 & y_2^2 \mathrm{logit}(y_2(\tilde{x}^{(2)})^T \tilde{\beta}) \mathrm{logit}(-y_2(\tilde{x}^{(2)})^T \tilde{\beta}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_N^2 \mathrm{logit}(y_N(\tilde{x}^{(N)})^T \tilde{\beta}) \mathrm{logit}(-y_N(\tilde{x}^{(N)})^T \tilde{\beta}) \end{pmatrix}$$

Since $y_i \in \{1, -1\}$, so $y_i^2 = 1$ and $\operatorname{logit}(y_i(\tilde{x}^{(1)})^T\tilde{\beta})\operatorname{logit}(-y_i(\tilde{x}^{(1)})^T\tilde{\beta}) = \operatorname{logit}((\tilde{x}^{(1)})^T\tilde{\beta})\operatorname{logit}(-(\tilde{x}^{(1)})^T\tilde{\beta})$ for all i.

$$D(h(f(\tilde{\beta})) = \begin{pmatrix} \operatorname{logit}((\tilde{x}^{(1)})^T\tilde{\beta})\operatorname{logit}(-(\tilde{x}^{(1)})^T\tilde{\beta}) & 0 & \cdots & 0 \\ 0 & \operatorname{logit}((\tilde{x}^{(2)})^T\tilde{\beta})\operatorname{logit}(-(\tilde{x}^{(2)})^T\tilde{\beta}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \operatorname{logit}((\tilde{x}^{(N)})^T\tilde{\beta})\operatorname{logit}(-(\tilde{x}^{(N)})^T\tilde{\beta}) \end{pmatrix}$$

$$\begin{aligned} \operatorname{Since} G(\mathbf{q}) &= \begin{pmatrix} \operatorname{logit}(q_1) \operatorname{logit}(-q_1) \\ \operatorname{logit}(q_2) \operatorname{logit}(-q_2) \\ \vdots \\ \operatorname{logit}(q_N) \operatorname{logit}(-q_N) \end{pmatrix} \text{ and } \operatorname{diag}(d) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}, \text{ so } \\ &\vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}, \text{ so } \\ &\operatorname{diag}(G(f(\tilde{\beta}))) &= \begin{pmatrix} \operatorname{logit}((\tilde{x}^{(1)})^T \tilde{\beta}) \operatorname{logit}(-(\tilde{x}^{(1)})^T \tilde{\beta}) & 0 & \cdots & 0 \\ 0 & \operatorname{logit}((\tilde{x}^{(2)})^T \tilde{\beta}) \operatorname{logit}(-(\tilde{x}^{(2)})^T \tilde{\beta}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \operatorname{logit}((\tilde{x}^{(N)})^T \tilde{\beta}) \operatorname{logit}(-(\tilde{x}^{(N)})^T \tilde{\beta}) \end{pmatrix} \\ &= D(h(f(\tilde{\beta}))) \end{aligned}$$

Therefore,

$$abla^2 \ell(ilde{eta}) = X^T \mathrm{diag}(G(f(ilde{eta}))) X$$

Part (d)

Take a $\mathbf{v} \in \mathbb{R}^{n+1}$, then we have

$$\mathbf{v}^T
abla^2 \ell(ilde{eta}) \mathbf{v} = \mathbf{v}^T X^T \mathrm{diag}(G(f(ilde{eta}))) X \mathbf{v} = \mathbf{V}^T \mathrm{diag}(G(f(ilde{eta}))) \mathbf{V} ext{ where } \mathbf{V} = X \mathbf{v}$$

Since this is in a sum form like $\sum_{i=1}^N d_i V_i^2$ where d_i is the ith diagonal element in $\mathrm{diag}(G(f(\tilde{\beta})))$ and V_i is the ith element in \mathbf{V} . We know that $\mathrm{logit}(x)>0$ for all $x\in\mathbb{R}$ and $V_i^2\geq 0$, so $d_iV_i^2\geq 0$ for all i. Thus, $\nabla^2\ell(\tilde{\beta})$ is positive semidefinite.

Part (e)

From part (d), we know that $d_i>0$ for all i. Then we have to make $V_i>0$ for all i. It means that $V\neq 0$ if $v\neq 0$. Since V=Xv, we just need that the null space of X is a set contains only ${\bf 0}$ vector. Therefore, X needs to have full rank to make sure that $\nabla^2\ell(\tilde{\beta})$ is positive definite.

Part (f)

Part (d) and Part (e) imply the minimization program is always convex and the program is strictly convex if X has full rank. Thus, all critical points are solutions to this program. When X has full rank, there will be a unique solution to this program.