HOMEWORK PROBLEMS 06, ANLY 561, FALL 2017
DUE 10/27/17

Exercises:

1. In multiple logistic regression, we have data $\{(\mathbf{x}^{(i)},\, y_i)\}_{i=1}^N$ where $\mathbf{x}^{(i)} \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$ (note that we are using $y = -1$ instead of $y = 0$ for simplicity). We model the probability that y equals 1 given $\mathbf{x}$ and parameters $\widetilde{\beta} \in \mathbb{R}^{n+1}$ by

$$\text{Prob}(y = 1|\mathbf{x}; \widetilde{\beta}) = \text{logit}(\widetilde{\mathbf{x}}^T \widetilde{\beta}),$$

and therefore

$$\text{Prob}(y = -1|\mathbf{x}; \widetilde{\beta}) = \text{logit}(-\widetilde{\mathbf{x}}^T \widetilde{\beta}),$$

where

$$\widetilde{\mathbf{x}}^{(i)} = \begin{pmatrix} 1 \\ \mathbf{x}^{(i)} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,n} \end{pmatrix} \text{ and } \widetilde{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}.$$

Note that $\widetilde{\mathbf{x}}^{(i)}, \widetilde{\beta} \in \mathbb{R}^{n+1}$. The **likelihood** of $\widetilde{\beta}$ given the data $\{(\mathbf{x}^{(i)},\, y_i)\}_{i=1}^N$ is then

$$\mathcal{L}(\widetilde{\beta}) = \prod_{k=1}^{N} \text{logit}(y_i \cdot (\widetilde{\mathbf{x}}^{(i)})^T \widetilde{\beta}),$$

and hence the negative log likelihood is

$$\ell(\widetilde{\beta}) = -\sum_{k=1}^{N} \log \text{logit}(y_i \cdot (\widetilde{\mathbf{x}}^{(i)})^T \widetilde{\beta}).$$

(a) Show that $\ell(\widetilde{\beta}) = s(g(f(\widetilde{\beta})))$ where $f : \mathbb{R}^n \to \mathbb{R}^N$, $g : \mathbb{R}^N \to \mathbb{R}^N$, and $s : \mathbb{R}^N \to \mathbb{R}$ are given by

$$f(\widetilde{\beta}) = X\widetilde{\beta} \text{ for } X = \begin{pmatrix} (\widetilde{\mathbf{x}}^{(1)})^T \\ (\widetilde{\mathbf{x}}^{(2)})^T \\ \vdots \\ (\widetilde{\mathbf{x}}^{(N)})^T \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & x_{N,2} & \cdots & x_{N,n} \end{pmatrix},$$

$$g(\mathbf{q}) = \begin{pmatrix} -\log \text{logit}(y_1 q_1) \\ -\log \text{logit}(y_2 q_2) \\ \vdots \\ -\log \text{logit}(y_N q_N) \end{pmatrix},$$

and

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

where $\mathbf{1}_N \in \mathbb{R}^N$ is the vector with entries all equal to 1.

(b) Use the multivariate chain rules and the product rule to show that

$$\nabla \ell(\widetilde{\beta}) = X^T h(f(\widetilde{\beta})))$$

where f is the function defined in part (a), $h : \mathbb{R}^N \to \mathbb{R}^N$ is

$$h(\mathbf{q}) = \begin{pmatrix} -y_1 \text{logit}(-y_1 q_1) \\ -y_2 \text{logit}(-y_2 q_2) \\ \vdots \\ -y_N \text{logit}(-y_N q_N) \end{pmatrix}.$$

(c) Use the multivariate chain rules and the product rule to show that

$$\nabla^2 \ell(\widetilde{\beta}) = X^T \mathrm{diag}(G(f(\widetilde{\beta})))X$$

where $G : \mathbb{R}^N \to \mathbb{R}^N$ is

$$G(\mathbf{q}) = \begin{pmatrix} \mathrm{logit}(q_1)\mathrm{logit}(-q_1) \\ \mathrm{logit}(q_2)\mathrm{logit}(-q_2) \\ \vdots \\ \mathrm{logit}(q_N)\mathrm{logit}(-q_N) \end{pmatrix},$$

and $\mathrm{diag} : \mathbb{R}^N \to M_{N,N}$ is

$$\mathrm{diag}(d) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}.$$

In other words, $\mathrm{diag}(d)$ is the diagonal matrix with diagonal entries given by the entries of d.

(d) Why is $\nabla^2 \ell(\widetilde{\beta})$ always positive semidefinite?

(e) When is $\nabla^2 \ell(\widetilde{\beta})$ always positive definite?

(f) What do parts (d) and (e) imply about the program

$$\min_{\widetilde{\beta} \in \mathbb{R}^n} \ell(\widetilde{\beta})?$$