[STAT409] Homework 4

Given a set of data $(y_i, \mathbf{x}_i) \in \{0,1\} \times \mathbb{R}^p, i = 1, \cdots, n$

1. Logistic Regression (LR) assumes

$$y_i \mid \mathbf{x}_i \sim \text{Bernoulli}(p_\beta(\mathbf{x}_i))$$

where

$$\log\left\{\frac{p_\beta(\mathbf{x}_i)}{1 - p_\beta(\mathbf{x}_i; \boldsymbol{\beta})}\right\} = \boldsymbol{\beta}^T\mathbf{x}_i \quad \text{or equivalently} \quad p_\beta(\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T\mathbf{x}_i)},$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ and $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^T$.

(a) Show that the log-likelihood of the logistic regression model is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left[y_i(\boldsymbol{\beta}^T\mathbf{x}_i) - \log\{1 + \exp(\boldsymbol{\beta}^T\mathbf{x}_i)\}\right]$$

(b) Show that the gradient vector $\nabla\ell(\boldsymbol{\beta})$ and Hessian matrix $\mathbf{H}(\boldsymbol{\beta})$ of $\ell(\boldsymbol{\beta})$ are respectively given by

$$\nabla\ell(\boldsymbol{\beta}) := \frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \mathbf{X}^T(\mathbf{y} - \mathbf{p}(\boldsymbol{\beta})) \quad \text{and} \quad \mathbf{H}(\boldsymbol{\beta}) := \frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = -\mathbf{X}^T\mathbf{W}\mathbf{X}$$

where

$$\mathbf{p}(\boldsymbol{\beta}) = (p_{\boldsymbol{\beta}}(\mathbf{x}_1), \cdots, p_{\boldsymbol{\beta}}(\mathbf{x}_n))^T, \quad \text{and} \quad \mathbf{W}(\boldsymbol{\beta}) = \text{diag}\{\mathbf{p}(\boldsymbol{\beta})\{1 - \mathbf{p}(\boldsymbol{\beta})\}\}$$

with $\mathbf{y} = (y_1, \cdots, y_n)^T$ and $\mathbf{X}$ being $n \times p$ design matrix whose (i,j)th element is $x_{ij}, i = 1, \cdots, n; j = 1, \cdots, p$.

2. Linear Support Vector Machine that seeks an optimal separating hyperplane $f(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T\mathbf{x}$ solves

$$\begin{aligned} &\min_{\beta_0, \boldsymbol{\beta}, \xi_i} \boldsymbol{\beta}^T\boldsymbol{\beta} + C\sum_{i=1}^n \xi_i \\ &\text{subject to} \;\; y_i(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i) \geq 1 - \xi_i, \quad i = 1, \cdots, n \\ &\qquad\qquad\quad \xi_i \geq 0, \qquad\qquad\qquad\quad i = 1, \cdots, n. \end{aligned}$$

(a) Introducing Lagrange multipliers $\alpha_i, \gamma_i \geq 0, i = 1, \cdots, n$, the primal function of the above is given by

$$\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\beta} + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i\{1 - y_i(\beta_0 - \boldsymbol{\beta}^T\mathbf{x}_i) - \xi_i\} - \gamma_i\sum_{i=1}^n \xi_i$$

Derive stationary conditions given in the lecture note.

(b) Show that corresponding dual problem is given by

$$\begin{aligned} &\max_{\alpha_1, \cdots, \alpha_n} \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \\ &\text{subject to } 0 \leq \alpha_i \leq C, \quad i = 1, \cdots, n; \\ &\qquad \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned}$$