DS-GA 3001.009 Applied Statistics: Homework #3

Due on Thursday, October 5, 2023

Please hand in your homework via Gradescope (entry code: RKXJN2) before 11:59 PM.

1. Let $p_\theta(y) = \exp(\theta \cdot T(y) - A(\theta))h(y)$ be a general one-dimensional exponential family, and consider a GLM $y_i \sim p_{\theta_i}$ with $\theta_i = \langle \beta, x_i \rangle$ for all $1 \le i \le n$.

   (a) Let $\widehat{\beta}$ be the MLE of β based on $(x_1, y_1), \cdots, (x_n, y_n)$. Prove Hoeffding's formula:

   $$D_+(\widehat{\beta}; \beta) = 2 \log \frac{p_{\widehat{\beta}}(y_1, \cdots, y_n \mid x_1, \cdots, x_n)}{p_\beta(y_1, \cdots, y_n \mid x_1, \cdots, x_n)}$$

   holds for any β, where $D_+(\cdot;\cdot)$ denotes the total deviance.

   (b) Let $\widehat{\beta}_0$ be the MLE restricted to a subset $\beta \in \Omega_0 \subseteq \mathbb{R}^p$; assume that Hoeffding's formula also holds for $\widehat{\beta}_0$ whenever $\beta \in \Omega_0$. Prove the deviance additivity theorem:

   $$D_+(\widehat{\beta}; \widehat{\beta}_0) = D_+(\widehat{\beta}; \beta) - D_+(\widehat{\beta}_0; \beta), \quad \forall \beta \in \Omega_0.$$

2. Let $\pi = (\pi_1, \cdots, \pi_k)$ be a probability vector, i.e. $\pi_j \ge 0$ for all $j = 1, \cdots, k$, $\sum_{j=1}^k \pi_j = 1$. Let p_π denote the statistical model $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \pi$ with sample size n.

   (a) Write out the log-likelihood $\ell_\pi(Y_1, \cdots, Y_n) = \log p_\pi(Y_1, \cdots, Y_n)$.

   (b) Let $(\pi_1, \cdots, \pi_{k-1})$ be the free parameters, and $\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j$. Show that the score function $s_\pi = (s_{\pi,1}, \cdots, s_{\pi,k-1})$ is given by

   $$s_{\pi,j}(Y_1, \cdots, Y_n) = \sum_{i=1}^n \left(\frac{\mathbb{1}(Y_i = j)}{\pi_j} - \frac{\mathbb{1}(Y_i = k)}{\pi_k} \right).$$

   (c) Verify that the Fisher information matrix $I(\pi)$ is given by

   $$I(\pi) = n \left(\operatorname{diag}(\pi_1^{-1}, \cdots, \pi_{k-1}^{-1}) + \frac{\mathbf{1}\mathbf{1}^\top}{\pi_k} \right),$$

   where $\mathbf{1} \in \mathbb{R}^{k-1}$ is the column vector consisting of all ones.

   (d) Using the Woodbury matrix identity (consult wikipedia), compute $I(\pi)^{-1}$. Compare your result with your answer to 2(a) in HW2. What do you find?

3. Coding: we will implement Lindsey's method for density estimation. Given $z_1, \cdots, z_{200} \sim p_Z$ (in the experiment we set $p_Z = \mathcal{N}(0.5, 1)$), we aim to fit p_Z using

   $$p_\theta(z) \propto \exp\left(\sum_{j=1}^5 \theta_j z^j \right) h(z)$$

   with $h(z) = \exp(-z^2/2)$. In other words, the fitted exponent is a degree-5 polynomial of z. In this problem, we will:

(a) use Lindsey's method to fit a full model $\theta \in \mathbb{R}^5$;

(b) use model selection techniques (AIC and Lasso) to fit a reduced model.

Based on the inline instructions, fill in the missing codes in `https://tinyurl.com/mr34wr63`. Be sure to submit a pdf with your codes, outputs, and colab link.

4. (Bonus question, 5 pts) In this problem we show that the map

$$(x, y) \mapsto g(x, y) = \log \left(\frac{1}{1 + e^{-x}} - \frac{1}{1 + e^{-y}} \right), \quad x, y \in \mathbb{R}, x \geq y$$

is concave, which implies the concavity of the MLE objective in the ordered logit model. To this end we use the following Prékopa-Leindler inequality.

Theorem 1 (Prékopa-Leindler)**.** *If $(u, v) \mapsto f(u, v) \in [0, \infty)$ is log-concave for $u \in \mathbb{R}^m, v \in \mathbb{R}^n$, the partial integration $u \mapsto \int_{\mathbb{R}^n} f(u, v) dv$ is also log-concave.*

(a) For $x \geq y, t \in \mathbb{R}$, show that

$$f(x, y, t) = \frac{e^t}{(1 + e^t)^2} \mathbb{1}(y \leq t \leq x)$$

is log-concave in (x, y, t).

(b) Use Prékopa-Leindler to conclude that $g(x, y)$ is concave in (x, y).

(c) Use the above program to prove that $(x, y) \mapsto \log(\Phi(x) - \Phi(y))$ is jointly concave in $(x, y) \in \mathbb{R}^2$ with $x \geq y$, where Φ is the CDF of the standard normal distribution. Choosing $y \to -\infty$, this gives an alternative proof that $x \mapsto \log \Phi(x)$ is concave.