DS-GA 3001.001 Applied Statistics: Homework #2

Due on Thursday, October 10, 2024

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

1. Let $\pi = (\pi_1, \cdots, \pi_k)$ be a probability vector, i.e. $\pi_j \geq 0$ for all $j = 1, \cdots, k$, $\sum_{j=1}^{k} \pi_j = 1$. Let p_π denote the statistical model $Y \sim \pi$, i.e. $p_\pi(Y = j) = \pi_j$ for all $j = 1, \ldots, k$.

   (a) Write out the log-likelihood $\ell_\pi(Y) = \log p_\pi(Y)$.

   (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 $\dot{\ell}_\pi = (\dot{\ell}_{\pi,1}, \cdots, \dot{\ell}_{\pi,k-1})$ is given by

   $$\dot{\ell}_{\pi,j}(Y) = \frac{\mathbb{1}(Y = j)}{\pi_j} - \frac{\mathbb{1}(Y = k)}{\pi_k}.$$

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

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

   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 3(a) in HW1. What do you find?

2. A dataset consists of n observations $(x_1, y_1), \cdots, (x_n, y_n)$, with $x_i \in \mathbb{R}^p, y_i \in \mathbb{N}$, following a multinomial model $(y_1, \cdots, y_n) \sim \mathrm{Multi}(N; (p_1, \cdots, p_n))$ with

   $$p_i = \frac{\exp(x_i^\top \beta)}{\sum_{j=1}^{n} \exp(x_j^\top \beta)}.$$

   (a) Show that the log-likelihood under this model is given by $\ell_{\mathrm{M}}(\beta) + c$, where

   $$\ell_{\mathrm{M}}(\beta) = \sum_{i=1}^{n} y_i \left(x_i^\top \beta - \log \left(\sum_{j=1}^{n} \exp(x_j^\top \beta) \right) \right),$$

   and $c \in \mathbb{R}$ is independent of β.

   (b) The Poissonization trick introduces an additional parameter $\phi \in \mathbb{R}$ and the following log-likelihood

   $$\ell_{\mathrm{P}}(\beta, \phi) = \sum_{i=1}^{n} \left(y_i(x_i^\top \beta + \phi) - e^{x_i^\top \beta + \phi} \right).$$

   Show that ℓ_{M} is the profile likelihood of ℓ_{P}, i.e. $\ell_{\mathrm{M}}(\beta) = \max_{\phi \in \mathbb{R}} \ell_{\mathrm{P}}(\beta, \phi) + c'$ for some constant $c' \in \mathbb{R}$ independent of β.

   (c) How does the result in (b) justify the use of Poissonization in Lindsey's method? You may assume $\Delta_k \equiv \Delta$ and $h(z_k) \equiv 1$ in your discussion.

3. In class we talked about how to estimate β in the Cox model. This problem investigates the estimation of the baseline survival function $S(t)$ (i.e. the survival function for an individual with $x = 0$).

   (a) Based on the lecture note, explain why the following is a reasonable estimator:

   $$\widehat{S}(t) = \exp\left(-\sum_{i:t_i \le t} \frac{\mathbb{1}(\Delta_i = 1)}{\sum_{k \in R_i} \exp(x_k^\top \widehat{\beta})}\right).$$

   Here R_i is the risk set at time t_i, and $\widehat{\beta}$ is the estimate of β from the Cox model.

   (b) If there is no feature (i.e. $\beta = \widehat{\beta} = 0$), comment on the similarities and differences between the above estimator and the Kaplan-Meier estimator for $S(t)$.

4. Coding I: 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.

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

5. Coding II: we will explore an AIDS dataset and understand the effects of different treatments on the survival curves for different patients. Based on the inline instructions, fill in the missing codes in `https://tinyurl.com/4bdcyy7c`. Be sure to submit a pdf with your codes, outputs, and colab link.

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

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

   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 $x \ge y, t \in \mathbb{R}$, show that

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

   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.