Reporting assignment
統計数理

2016-5-30

Taiji Suzuki (Room W707, post W8-46)
e-mail: suzuki.t.ct@m.titech.ac.jp

Solve the following problems. Post the report to the mailbox located on 3rd floor between East and West wings of W8 building. You may answer in either Japanese or English.
Due date： 13th June (Monday).

1. Consider a linear model
$$Y = \mathbf{X}\beta^* + \epsilon$$
where $Y \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n\times p}$, and $\epsilon = (\epsilon_i)_{i=1}^n \in \mathbb{R}^n$. Suppose that ϵ_i is an i.i.d. noise such that $\mathrm{E}[\epsilon_i] = 0$ and $\mathrm{E}[\epsilon_i^2] = \sigma^2$. In this setting, evaluate the in-sample predictive error
$$\mathrm{E}_{Y|X}\left[\frac{1}{n}\|\mathbf{X}\beta^* - \mathbf{X}\hat{\beta}_{\mathrm{LS}}\|^2\right].$$
of the least squared estimator $\hat{\beta}_{\mathrm{LS}} \in \mathbb{R}^p$.
2. (Stein's shrinkage estimator) Let $X = [X_1, \dots, X_d]^\top \in \mathbb{R}^d$ be distributed from multivariate normal $N(\boldsymbol{\mu}, I)$ (mean $\boldsymbol{\mu}$ and variance-covariance I). Assume $d \geq 3$. Using Stein's identity, show that
$$\boldsymbol{\delta} = \left(1 - \frac{(d-2)}{\|X\|^2}\right) X$$
satisfies
$$\mathrm{E}_{X\sim N(\boldsymbol{\mu},I)}[\|X - \boldsymbol{\mu}\|^2] > \mathrm{E}_{X\sim N(\boldsymbol{\mu},I)}[\|\boldsymbol{\delta}(X) - \boldsymbol{\mu}\|^2] \quad (\forall \boldsymbol{\mu} \in \mathbb{R}^d).$$
3. Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function. Show that, for all $x, y \in \mathbb{R}^d$, and all $g \in \partial f(x)$ and $g' \in \partial f(y)$, it holds that
$$\langle x - y, g - g' \rangle \geq 0.$$
4. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function, $A \in \mathbb{R}^{q\times d}$ be a column full rank matrix and $b \in \mathbb{R}^q$. Suppose that, for all $u \in \mathbb{R}^q$, $\bar{f}(u) = \inf_{x\in\mathbb{R}^d}\{f(x) \mid u = b - Ax\} > -\infty$. Prove that $\bar{f}(u)$ is a convex function.
5. Let $f(u) = \log(1 + \exp(-u))$ and $h(\beta) = \|\beta\|_p^p$ $(p > 1)$. Derive the dual optimization problem of
$$\min_{\beta\in\mathbb{R}^d}\left\{\sum_{i=1}^n f(x_i^\top\beta) + h(\beta)\right\}$$
using Fenchel's duality theorem.
6. Derive the Moreau envelope of $f(x) = \|x\|_1$ $(x \in \mathbb{R}^d)$. Remind that the Moreau envelope $\tilde{f}$ of a closed convex function f is given by $\tilde{f} : q \mapsto \min_x\{f(x) + \frac{1}{2}\|x - q\|^2\}$.
7. Consider the Gaussian kernel $k(x, y) = \exp(-\frac{(x-y)^2}{2\sigma^2})$ $(x, y \in \mathbb{R})$. Derive its Fourier transform:
$$\phi(w) = \int e^{-\sqrt{-1}wz}\exp(-z^2/2\sigma^2)\mathrm{d}z \quad (w \in \mathbb{R}).$$
Then conclude that the Gaussian kernel is positive definite and characteristic (you may use the theorems that are presented in the lecture without proofs).

8. Suppose that $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ $(i = 1, \dots, n)$ where $\boldsymbol{\mu} \in \mathbb{R}^d$ is unknown and $\Sigma \in \mathbb{R}^{d \times d}$ is known. Now we want to estimate $\boldsymbol{\mu}$. Derive the posterior distribution of $\boldsymbol{\mu}$ based on the observation $x = \{\boldsymbol{x}_i\}_{i=1}^n$ when the prior distribution of that is $N(\boldsymbol{m}_0, S_0)$.
   Hint: remind that the posterior distribution for the prior π is given by
   $$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)\mathrm{d}\theta}.$$
   Substitute the Gaussian probability density function into the likelihood term $p(x|\theta)$ and the Gaussian prior into the prior term $\pi(\theta)$ for $\theta = \boldsymbol{\mu}$.
9. (optional) Implement a machine learning technique that is introduced in the lecture, and apply it to some data. Then, report the result. You may use any method, any programming language such as C, python, R, matlab, and any data you like.