Reporting assignment
統計数理 1

2014-7-22 (2015/2/24 修正)

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： August 8th (Friday).

1. (Stein's identity) Let $X = [X_1, \dots, X_d]^\top \in \mathbb{R}^d$ be distributed from multivariate normal $N(\boldsymbol{\mu}, \sigma^2 I)$ (mean $\boldsymbol{\mu}$ and variance-covariance $\sigma^2 I$), and $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$ ($X \mapsto \boldsymbol{f}(X) = [f_1(X), \dots, f_d(X)]^\top$). Assume $\mathrm{E}_X[f_i(X)]$ exists and f_i is differentiable almost everywhere for all i. Then show that
$$\mathrm{E}_X\left[-2\langle \boldsymbol{\mu} - X, \boldsymbol{f}(X)\rangle\right] = 2\sigma^2 \mathrm{E}_X\left[\sum_{i=1}^d \frac{\partial f_i(X)}{\partial X_i}\right].$$
2. (Stein's estimator) Suppose the same setting as problem 1 with $\sigma = 1$. Assume $d \geq 3$. Using Stein's identity, show that
$$\boldsymbol{\delta} = \left(1 - \frac{1}{\|X\|^2}\right) X$$
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. Let
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0.5 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$
Then find one of the optimal solutions of the following optimization problem:
$$\hat{x} \in \mathop{\arg\min}_{x \in \mathbb{R}^3} \frac{1}{2}\|Ax - y\|^2 + \|x\|_1.$$
4. Show that if a convex function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at x, then the subgradient ($\partial f(x)$) of f at x consists of only one element $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$.
5. Show that for all $\beta \in \mathbb{R}^d$
$$\|\beta\|_2 \leq \|\beta\|_1 \leq \sqrt{d}\|\beta\|_2,$$
where $\|x\|_p = (\sum_{i=1}^d |x_i|^p)^{\frac{1}{p}}$ $(1 \leq p < \infty)$.
6. Derive the convex conjugate function of ℓ_p-norm $h(\beta) = \|\beta\|_p$ ($\beta \in \mathbb{R}^d$) where $p \geq 1$.
7. Derive the convex conjugate function of the logistic loss $f(u) = \log(1 + \exp(-u))$ ($u \in \mathbb{R}$).
8. Show that the linear kernel $k(x, y) = x^\top y$ is *not* characteristic[*1].
9. Consider the Gaussian kernel $k(x, y) = \exp(-\frac{(x-y)^2}{2\sigma^2})$ ($x, y \in \mathbb{R}$). Compute its Fourier transform:
$$\phi(w) = \int e^{\sqrt{-1}wz} \exp(-z^2/2\sigma^2)\mathrm{d}z \quad (w \in \mathbb{R}).$$
It is known that, if $\phi(w) > 0$ ($\forall w \in \mathbb{R}$), then the kernel is characteristic.

(optional) Apply a machine learning technique to real data which you collected, and report the result. You may use any method, any programming language such as C, python, R, matlab, and any data you like.

[*1] 特性的