Reporting assignment

京都大学集中講義
「機械学習と深層学習の数理と応用」

2018-12-20

Taiji Suzuki
e-mail: taiji@mist.i.u-tokyo.ac.jp

Solve the following problems.
Due date: 10 January/2019.

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 $\mathrm{E}[\epsilon_i^2] = \sigma^2$ and $\mathbf{X}^\top\mathbf{X}$ is full rank. 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 $X = [X_1, \ldots, X_d]^\top \in \mathbb{R}^d$ be distributed from multivariate normal $N(\boldsymbol{\mu}, I)$ (mean $\boldsymbol{\mu}$ and variance-covariance I). Assume $d \geq 3$. Then, show that
$$\boldsymbol{\delta} = \left(1 - \frac{(d-2)}{\|X\|^2}\right) X$$
satisfies
$$\mathrm{E}_{X\sim N(\boldsymbol{\mu},I)}[\|X - \boldsymbol{\mu}\|^2] > \mathrm{E}_{X\sim N(\boldsymbol{\mu},I)}[\|\boldsymbol{\delta}(X) - \boldsymbol{\mu}\|^2] \quad (\forall \boldsymbol{\mu} \in \mathbb{R}^d).$$
You may use the following *Stein's identy* without proof: For a function $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$ ($X \mapsto \boldsymbol{f}(X) = [f_1(X), \ldots, f_d(X)]^\top$) such that $\mathrm{E}_X[f_i(X)]$ exists and f_i is differentiable almost everywhere for all i, it holds that
$$\mathrm{E}_X\left[-2\langle \boldsymbol{\mu} - X, \boldsymbol{f}(X)\rangle\right] = 2\sigma^2 \mathrm{E}_X\left[\sum_{i=1}^d \frac{\partial f_i(X)}{\partial X_i}\right].$$
3. For $1 \leq q < \infty$, let $\|w\|_q := (\sum_{i=1}^d |w_i|^q)^{1/q}$ for $w \in \mathbb{R}^d$, and $\mathcal{H}_q := \{f(x) = w^\top x \ (x \in \mathbb{R}^d) \mid \|w\|_q \leq 1, w \in \mathbb{R}^d\}$. Given $x_1, \ldots, x_n \in \mathbb{R}^n$, its empirical Rademacher complexity is denoted by
$$\hat{R}_n(\mathcal{H}_q) = \mathrm{E}_\sigma\left[\sup_{f\in\mathcal{H}_q} \frac{1}{n}\sum_{i=1}^n \sigma_i f(x_i) \;\middle|\; x_1, \ldots, x_d\right].$$
Now, suppose that $1 \leq p, q < \infty$ satisfies $q < p$. Then, show that
$$\hat{R}_n(\mathcal{H}_p) \leq d^{1/p^* - 1/q^*}\hat{R}_n(\mathcal{H}_q)$$
where $p^* = p/(p-1)$ and $q^* = q/(q-1)$.
4. Prove the Massart's theorem: For a finite set of functions $\mathcal{F} = \{f_1, \ldots, f_M\}$ where each f_i ($i = 1, \ldots, M$) is a function from $\mathbb{R}^d$ to $\mathbb{R}$ satisfying $\sup_x |f_i(x)| \leq R$, it holds that
$$\hat{R}_n(\mathcal{F}) \leq R\sqrt{\frac{2\log M}{n}}.$$
Hint: You may use the following inequalities:
   - Jensen's inequality: $\exp(s\mathrm{E}_\sigma[g(\sigma_1, \ldots, \sigma_n)]) \leq \mathrm{E}_\sigma[\exp(sg(\sigma_1, \ldots, \sigma_n))]$ for $s \in \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$.
   - $\exp(\max_{f\in\mathcal{F}} F(f)) \leq \sum_{f\in\mathcal{F}} \exp(F(f))$ for $F : \mathcal{F} \to \mathbb{R}$.
   - Hoeffding's inequality: $\mathrm{E}_{\sigma_i}[\exp(\sigma_i a)] \leq \exp(a^2/2)$ for $a \in \mathbb{R}$.
5. Suppose that $\|x\|_\infty < 1$ (a.s.). Derive an upper bound of the Rademacher complexity of a neural network model:
$$\mathcal{F} = \left\{\sum_{j=1}^M \alpha_j \eta(a_j^\top x) \mid \alpha_j \in \mathbb{R},\ a_j \in \mathbb{R}^d,\ \max_{1\leq j\leq M}\|a_j\|_p \leq C_1,\ \|\alpha\|_q \leq C_2\right\}$$
where $1 \leq p, q \leq \infty$ and $\eta(u) = \max\{u, 0\}$.