集中講義「深層学習および機械学習の数理」

レポート課題
締め切り：9 月 25 日

2020-9-2

Taiji Suzuki
6th Laboratory of Mathematical Informatics
(Room 352, Faculty of Engineering Building 6)
e-mail: taiji@mist.i.u-tokyo.ac.jp

以下の問題のうち 2 問以上を選択して解答してください．レポートは日本語で構いません．

1. Consider the second order polynomial kernel $k(x,y) = (1 + x^\top y)^2$ on $\mathbb{R}^2$. Construct a function having a form $f(x) = \sum_{j=1}^m \alpha_j k(x_j, x)$ with $x_j \in \mathbb{R}^2$ and $\alpha_j \in \mathbb{R}$ $(j = 1, \dots, m)$ such that
$$\begin{cases} f(x) \leq 0 & (\|x\| \leq 1), \\ f(x) > 0 & (\|x\| > 1), \end{cases}$$
for $x \in \mathbb{R}^2$
2. Derive the dual problem of support vector machine by using Fenchel's duality theorem:
$$\min_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^n \max\left\{1 - y_i \sum_{j=1}^n k(x_i, x_j)\alpha_j, 0\right\} + \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j).$$
Fenchel's duality theorem is given in Theorem 2.
3. Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a characteristic kernel that has $\mathcal{H}$ as its corresponding RKHS. For a probability measure P, define
$$m_P(x) = \int k(x, X) \mathrm{d}P(X).$$
Assume that k is transition invariant, i.e. there exits ψ such that $k(x,y) = \psi(x - y)$. Then show that, for any probability measures P and Q,
$$\|m_P - m_Q\|_{\mathcal{H}} = 0 \quad \Leftrightarrow \quad P = Q.$$
Hint: Use Bochner's theorem and the necessary and sufficient condition of positive definiteness.
4. 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 $x_1, \dots, x_n \in \mathbb{R}^n$, its empirical Rademacher complexity is denoted by
$$\hat{R}_n(\mathcal{H}_q) = \mathrm{E}_\sigma\left[\sup_{f \in \mathcal{H}_q} \left|\frac{1}{n}\sum_{i=1}^n \sigma_i f(x_i)\right| \ \middle|\ x_1, \dots, x_d\right].$$
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 $q^* = q/(q-1)$.
5. Prove the Massart's theorem: For a finite set of functions $\mathcal{F} = \{f_1, \dots, f_M\}$ where each f_i $(i = 1, \dots, M)$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$ satisfying $\sup_x |f_i(x)| \leq R$, it holds that
$$\hat{R}_n(\mathcal{F}) = \mathrm{E}_\sigma\left[\sup_{f \in \mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n \sigma_i f(x_i)\right| \ \middle|\ (x_i)_{i=1}^n\right] = \mathrm{E}_\sigma\left[\sup_{f \in \mathcal{F} \cup \{-f' \mid f' \in \mathcal{F}\}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(x_i) \ \middle|\ (x_i)_{i=1}^n\right]$$
$$\leq R\sqrt{\frac{2 \log 2M}{n}}.$$
Hint: You may use the following inequalities:

- Jensen's inequality: $\exp(s\mathrm{E}_\sigma[g(\sigma_1,\dots,\sigma_n)]) \leq \mathrm{E}_\sigma[\exp(sg(\sigma_1,\dots,\sigma_n))]$ 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 $a \in \mathbb{R}$.

6. 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 $\eta(u) = \max\{u, 0\}$.
7. Write your opinion about the future direction of machine learning and artificial intelligence techniques. Of course, you may discuss the future of deep learning.

Definition 1 (Legendre transform, convex conjugate) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Its convex conjugate function $f^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$f^*(y) = \sup_{x\in\mathbb{R}^d} \{x^\top y - f(x)\}.$$

This transformation $f \mapsto f^*$ is called *Legendre transformation.*

Theorem 2 (Fenchel's duality theorem (special case)) Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ be convex functions, and let $A \in \mathbb{R}^{n\times d}$. Then it holds that

$$\inf_{x\in\mathbb{R}^d} \{f(Ax) + g(x)\} = -\inf_{y\in\mathbb{R}^n} \{f^*(y) + g^*(-A^\top y)\}$$

where f^*, g^* are the convex conjugate functions of f and g respectively.