Sample Problems 5

To be discussed on 21.06.2024

Sample Problem 5.1: Rademacher Complexity

Let $\mathcal{X} = [0,1] \subset \mathbb{R}$ and denote by $\mathcal{D}$ the uniform distribution on $\mathcal{X}$. Define the linear function class

$$\mathcal{F} = \{f(x) = \langle v, x \rangle : \|v\|_2 \le \rho\}$$

Prove that

$$\mathcal{R}(\mathcal{D}, m) \le \frac{\rho\sqrt{m}}{m+1}$$

Sample Problem 5.2: Hard SVM vs. Soft SVM

Prove or disprove the following statement: There exists a choice of parameter $\lambda > 0$ such that the solution of Soft SVM with parameter λ is identical to the solution of Hard SVM for *every set of separable training data.*

Sample Problem 5.3: Generalisation in one-class SVM

One-class SVM is a method used for anomaly detection. Here the training set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\} \subset \mathcal{X}$ consists of only non-anomalous samples, and the one-class SVM learns a classifier that labels a small region, containing S, by $+1$ and everything else by -1. One-class SVMs are usually defined using kernels, but in this problem, we consider linear one-class SVM.

Assume $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^p \mid \|\mathbf{x}\| \le \rho\}$. Given the positive examples $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, linear one-class SVM returns a classifier $\widehat{h} = \mathrm{sign}(\mathbf{w}^\top \mathbf{x})$ whose parameters are solutions of the optimisation

$$\begin{aligned} &\underset{\mathbf{w}\in\mathbb{R}^p, \xi\in\mathbb{R}^m, \nu\in\mathbb{R}}{\text{minimise}} \; \|\mathbf{w}\|^2 + \frac{1}{\lambda \cdot m}\sum_{i=1}^{m}(\xi_i - \nu) \\ &\text{subject to } \mathbf{w}^\top \mathbf{x}_i \ge \nu - \xi_i, \; \xi_i \ge 0, \quad \text{for all } i = 1, \ldots, m \end{aligned} \tag{$*$}$$

The following sub-problems show that the above optimisation is a Tikhonov regularised loss minimisation (RLM) problem with a convex Lipschitz loss. This view of one-class SVM as Tikhonov RLM allows one to derive generalisation error bounds for this method.

1. One can rewrite the optimisation $(*)$ as an unconstrained optimisation by eliminating $\xi_1, \ldots, \xi_m$ to obtain
$$\operatorname*{minimise}_{\mathbf{w}\in\mathbb{R}^p, \nu\in\mathbb{R}} \frac{1}{m}\sum_{i=1}^{m} \ell_{x_i}(\mathbf{w}, \nu) + \lambda\|\mathbf{w}\|^2$$
where $\ell_{\mathbf{x}_i}(\mathbf{w}, \nu)$ is a function of $\mathbf{x}_i, \mathbf{w}, \nu$. Give the expression of $\ell_{\mathbf{x}}(\mathbf{w}, \nu)$.

2. Define $\theta = (\mathbf{w}, \nu) \in \mathbb{R}^{p+1}$. Rewrite $\ell_{\mathbf{x}}(\theta)$ in terms of the vector θ and show that $\ell_{\mathbf{x}}(\theta)$ is convex with respect to θ.

3. Show that $\ell_{\mathbf{x}}(\theta)$ is Lipschitz with Lipschitz constant bounded by $\rho' \leq 1 + \sqrt{1+\rho^2}$ (in fact, you can find a better Lipschitz constant).