Homework 10

36-465/665, Spring 2021

Due at 6 pm on **Tuesday, 13 April** 2021

Agenda: A key point about kernel machines. **PLEASE NOTE THE DUE DATE!!!**

1. *Empirical Rademacher complexity of kernel machines.* We have fixed a kernel $K(x, z)$, with corresponding basis functions $\phi_1, \phi_2, \ldots$, so $K(x, z) = \sum_{j=1}^{\infty} \phi_j(x)\phi_j(z)$. We have observed a set of training points $x_1, x_2, \ldots x_n$. The **kernel matrix K** has entries $K_{ij} = K(x_1, x_2)$. We are interested in the empirical Rademacher complexity of the class S_c of all kernel machines of the form $s(x) = \sum_{j=1}^{\infty} \beta_j \phi_j(x)$, where $\|\beta\| \leq c$, for some $c > 0$. We'll abbreviate this as $s(x) = \beta \cdot \phi(x)$.

   a. (7) Show that

   $$\hat{\mathcal{R}}_n(S_c) = \mathbb{E}\left[\frac{1}{n} \max_{s \in S_c} \sum_{i=1}^{n} \sigma_i s(x_i)\right]$$

   That is, show that the absolute value sign in the definition of $\hat{\mathcal{R}}$ can be dispensed with in this case. *Hint*: HW6Q1.

   b. (7) Show that

   $$\hat{\mathcal{R}}_n(S_c) = \mathbb{E}\left[\max_{s \in S_c} \beta \cdot \frac{1}{n} \sum_{i=1}^{n} \sigma_i \phi(x_i)\right]$$

   c. (5) Show that

   $$\hat{\mathcal{R}}_n(S_c) \leq c\mathbb{E}\left[\left\|\frac{1}{n} \sum_{i=1}^{n} \sigma_i \phi(x_i)\right\|\right]$$

   Hint: For any kind of inner product, $(a \cdot b)^2 \leq (a \cdot a)^2 (b \cdot b)^2 = \|a\|^2 \|b\|^2$ (the Cauchy-Schwarz inequality, which you can use without proving).

   d. (7) Show that

   $$\hat{\mathcal{R}}_n(S_c) \leq \frac{c}{n} \sqrt{\mathbb{E}\left[\left(\sum_{i=1}^{n} \sigma_i \phi(x_i)\right) \cdot \left(\sum_{i=1}^{n} \sigma_i \phi(x_i)\right)\right]}$$

   Hint: $-u^{1/2}$ is a convex function of u.

   e. (8) Show that

   $$\hat{\mathcal{R}}_n(S_c) \leq \frac{c}{n} \sqrt{\mathbb{E}\left[\sum_{i=1}^{n} \phi(x_i) \cdot \phi(x_i)\right]}$$

   Hint: $\sigma^2 = 1$ (why?); what's $\mathbb{E}[\sigma_i \sigma_j]$ when $j \neq i$?

   f. (6) Show that

   $$\hat{\mathcal{R}}_n(S_c) \leq \frac{c}{n} \sqrt{\mathbb{E}\left[\sum_{i=1}^{n} K(x_i, x_i)\right]}$$

g. (7) Show that

$$\hat{\mathcal{R}}_n(S_c) \leq \frac{c}{n}\sqrt{\operatorname{tr}\mathbf{K}}$$

h. (8) Suppose we know that $\mathbb{P}\left(K(X,X) \leq r^2\right) = 1$. Show that

$$\mathcal{R}_n(S_c) \leq \frac{cr}{\sqrt{n}}$$

2. *Bounded coefficient vectors versus bound weights on points* The slides begin by saying kernel machines take the form $s(x) = \sum_{i=1}^{n} K(x, x_i)$, and that this is equivalent to functions of the form $\sum_{j=1}^{\infty} \beta_j \phi_j(x)$. The result about empirical Rademacher complexity in Q1 is in terms of β, but it's often more convenient to have a bound in terms of α. In what follows, we show how $\|\beta\|$ and $\|\alpha\|$ are related (even though those are vectors of different dimensions).

a. (6) Show that

$$\beta_j = \sum_{i=1}^{n} \alpha_i \phi_j(x_i)$$

b. (6) Show that

$$\|\beta\|^2 = \sum_{i=1}^{n} \sum_{i'=1}^{n} \alpha_i \alpha_{i'} \sum_{j=1}^{\infty} \phi_j(x_i)\phi_j(x_{i'})$$

c. (6) Show that

$$\|\beta\|^2 = \sum_{i=1}^{n} \sum_{i'=1}^{n} \alpha_i \alpha_{i'} K(x_i, x_{i'})$$

d. (6) Show that

$$\|\beta\|^2 = \alpha \cdot \mathbf{K}\alpha$$

e. (10) Consider the class of all kernel machines where $\alpha \cdot \mathbf{K}\alpha \leq c^2$. Could you check whether a given kernel machine was in this class, using the data and the kernel function alone? (That is, not knowing the data-generating distribution?) Use Q1 to state a bound on its empirical Rademacher complexity.

3. (1) How much time did you spend on this problem set?

Presentation rubric (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision.