Mehryar Mohri
Foundations of Machine Learning 2023
Courant Institute of Mathematical Sciences
Homework assignment 2
October 6, 2023
Due: October 31, 2023

A. Radmacher complexity

1. Consider the class of functions $\mathcal{H}$ mapping from $\mathbb{R}$ to $\{+1, -1\}$ such that

$$h(x) = \begin{cases} +1 \text{ for } x \in [a, b], \\ -1 \text{ otherwise }, \end{cases}$$

for some $a, b \in \mathbb{R}$. Use Sauer's lemma to give an upper bound on the growth function $\Pi_{\mathcal{H}}(m)$ and prove that the upper bound is tight in this example. Use it to derive an upper bound on $\mathfrak{R}_m(\mathcal{H})$.

2. Prove that for any $\alpha, \beta \in \mathbb{R}$ and any two hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$ of functions mapping from $\mathcal{X}$ to $\mathbb{R}$, the equality $\mathfrak{R}_m(\alpha\mathcal{H}_1 + \beta\mathcal{H}_2) = |\alpha|\mathfrak{R}_m(\mathcal{H}_1) + |\beta|\mathfrak{R}_m(\mathcal{H}_2)$ holds, where the linear combination of the two hypothesis sets are defined by $\alpha\mathcal{H}_1 + \beta\mathcal{H}_2 = \{\alpha h_1 + \beta h_2 : h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$.

3. Prove that if for two hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$ the inclusion $\mathcal{H}_1 \subseteq \mathcal{H}_2$ holds, then the following inequality holds for any finite sample S: $\widehat{\mathfrak{R}}_S(\mathcal{H}_1) \leq \widehat{\mathfrak{R}}_S(\mathcal{H}_2)$.

4. Let $\mathcal{H}_1$ be a family of functions mapping from $\mathcal{X}$ to $\{0, 1\}$ and let $\mathcal{H}_2$ be a family of functions mapping from $\mathcal{X}$ to $\{-1, +1\}$. Let $\mathcal{H} = \{h_1 h_2 : h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$. Show that the empirical Rademacher complexity of $\mathcal{H}$ for any sample S of size m can be bounded as follows:

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \leq \widehat{\mathfrak{R}}_S(\mathcal{H}_1) + \widehat{\mathfrak{R}}_S(\mathcal{H}_2).$$

[hint: write $h_1 h_2$ in a way such that you can apply Talagrand's lemma.]

B. VC-dimension

1. What is the VC-dimension of axis-aligned squares in $\mathbb{R}^2$? Is this value the same as the VC-dimension of squares (not necessarily axis-aligned) in $\mathbb{R}^2$? Why?

2. What is the VC-dimension of intersections of 2 axis-aligned squares in $\mathbb{R}^2$?

3. (a) For two concept classes $\mathcal{C}_1, \mathcal{C}_2$, define the concept class $\mathcal{C}$ by

$$\mathcal{C} = \{c_1 c_2 \mid c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2\}.$$

Prove that the following inequality holds:

$$\Pi_{\mathcal{C}}(m) \leq \Pi_{\mathcal{C}_1}(m)\Pi_{\mathcal{C}_2}(m).$$

   (b) Let $\mathcal{C}$ be a concept class whose VC-dimension is 3. Show that the VC-dimension of intersections of k concepts from $\mathcal{C}$ is upper bounded by $6k \log_2(3k)$. [hint: use Sauer's lemma and the result of (a).]

C. Support Vector Machines

1. (a) SVMs are "sparse" in the sense that the number of support vectors is usually small compared to total number of observations. Suppose we explicitly maximize sparsity by penalizing the L_2 norm of the vector $\boldsymbol{\alpha}$ that defines the weight vector $\mathbf{w}$:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, b, \boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{\alpha}\|^2 + C\left(\sum_{i=1}^{m} \xi_i\right) \\ \text{subject to} \quad & y_i\left(\left(\sum_{j=1}^{m} \alpha_j y_j \mathbf{x}_j\right) \cdot \mathbf{x}_i + b\right) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \alpha_i \geq 0, i \in [m]. \end{aligned} \tag{1}$$

   Show that the problem coincides with an instance of the primal optimization problem of SVMs, modulo the non-negativity constraint on $\boldsymbol{\alpha}$. You should indicate exactly how to view it as such.

   (b) Derive the dual optimization problem of (1).

2. Suppose we replace in the primal optimization problem of SVMs the penalty term $\sum_{i=1}^{m} \xi_i = \|\boldsymbol{\xi}\|_1$ with $\|\boldsymbol{\xi}\|_\infty = \max_{i=1}^{m} \xi_i$. Give the associated dual optimization problem. Show that it differs from the standard dual optimization problem of SVMs only by the constraints, which can be expressed in terms of $\|\boldsymbol{\alpha}\|_1$.