Assignment 5

00103335: Deep Learning and Reinforcement Learning

Fall 2026

Source. FML refers to M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, second edition (2018).

Give complete proofs. Results explicitly supplied as tools may be used without proof. All logarithms are natural unless a base is specified.

Problem 1. VC dimension (FML Exercises 3.13, 3.15, 3.17, 3.19, 3.20)

(a) **Unions of intervals (3.13).** Let $k \geq 1$, and let $\mathcal{I}_k$ be the family of all subsets of $\mathbb{R}$ that are unions of at most k closed bounded intervals, including the empty set. Prove that $\mathrm{VCdim}(\mathcal{I}_k) = 2k$.

(b) **A one-parameter family (3.15).** Consider $\{I_\alpha : \alpha \in \mathbb{R}\}$, where

$$I_\alpha = [\alpha, \alpha + 1] \cup [\alpha + 2, \infty).$$

Prove that $\mathrm{VCdim}(\{I_\alpha : \alpha \in \mathbb{R}\}) = 3$. *Hint:* for fixed sample points, track how their labels change as α passes the values $x_i - 2$, $x_i - 1$, and x_i. Include the labels at the critical values themselves.

(c) **Closed balls (3.17).** For $p \geq 1$, let

$$\mathcal{B}_p = \big\{\{x \in \mathbb{R}^p : \|x - c\|_2 \leq r\} : c \in \mathbb{R}^p,\ r \geq 0\big\}.$$

Prove that $\mathrm{VCdim}(\mathcal{B}_p) \leq p + 2$. *Hint:* lift x to $(x, \|x\|_2^2)$ and use the VC dimension of affine halfspaces.

(d) **A vector space of functions (3.19).** Let $\mathcal{F}$ be a vector space of real-valued functions on $\mathbb{R}^p$, with $\dim \mathcal{F} = r < \infty$. Prove that

$$\mathrm{VCdim}\big(\{\{x : f(x) \geq 0\} : f \in \mathcal{F}\}\big) \leq r.$$

Hint: for $r + 1$ arbitrary distinct points, consider the linear map $f \mapsto (f(x_1), \ldots, f(x_{r+1}))$.

(e) **Sine classifiers (3.20).** Define the binary class

$$\mathcal{H}_{\sin} = \{h_\omega : \omega \in \mathbb{R}\}, \qquad h_\omega(x) = \mathbf{1}_{\{\sin(\omega x) \geq 0\}}.$$

(i) For $x \neq 0$, put $A_x = \{x, 2x, 3x, 4x\}$. Prove that $\mathrm{VCdim}(\mathcal{H}_{\sin}|_{A_x}) \leq 3$, where $\mathcal{H}_{\sin}|_{A_x} = \{h|_{A_x} : h \in \mathcal{H}_{\sin}\}$ is the class restricted to A_x. For $x = 0$, prove that the restricted class has VC dimension 0.

(ii) Prove that $\mathrm{VCdim}(\mathcal{H}_{\sin}) = \infty$. *Hint:* show that $\{2^{-i} : 1 \leq i \leq m\}$ can be shattered for every $m \geq 1$; use binary digits.

Problem 2. Union of hypothesis classes (FML Exercise 3.24)

Let $\mathcal{A}$ and $\mathcal{B}$ be nonempty classes of functions from $\mathcal{X}$ to $\{0,1\}$, with finite VC dimensions

$$d_A = \mathrm{VCdim}(\mathcal{A}), \qquad d_B = \mathrm{VCdim}(\mathcal{B}).$$

Set $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$. For a class $\mathcal{H}$, define its growth function by

$$\Pi_{\mathcal{H}}(m) = \sup_{(x_1,\ldots,x_m)\in\mathcal{X}^m} \left|\{(h(x_1),\ldots,h(x_m)) : h \in \mathcal{H}\}\right|.$$

(a) Prove that, for every integer $m \geq 1$,

$$\Pi_{\mathcal{C}}(m) \leq \Pi_{\mathcal{A}}(m) + \Pi_{\mathcal{B}}(m).$$

(b) Use Sauer–Shelah's lemma to prove that, for every integer $m \geq d_A + d_B + 2$,

$$\Pi_{\mathcal{C}}(m) < 2^m.$$

Deduce that $\mathrm{VCdim}(\mathcal{C}) \leq d_A + d_B + 1$. *Hint:* use the binomial-sum form of Sauer–Shelah's lemma and the identity $\binom{m}{j} = \binom{m}{m-j}$.

Problem 3. Rademacher identities (FML Exercise 3.8)

Convention for Problems 3 and 4. These two problems use FML's convention *without an absolute value* inside the supremum:

$$\widehat{\mathcal{R}}_S(\mathcal{H}) := \mathbb{E}_\epsilon \sup_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \epsilon_i h(x_i), \qquad \mathcal{R}_m(\mathcal{H}) := \mathbb{E}_{S\sim\mu^m}\widehat{\mathcal{R}}_S(\mathcal{H}).$$

Here $S = (x_1,\ldots,x_m)$, $m \geq 1$, and μ is a probability distribution on the input space $\mathcal{X}$. The signs are independent of the sample. This differs from the absolute-value convention in Lecture 3. Assume all classes are nonempty and all quantities below are measurable and finite.

For $\alpha \in \mathbb{R}$ and classes $\mathcal{H}, \mathcal{H}'$ of real-valued functions, define

$$\alpha\mathcal{H} = \{\alpha h : h \in \mathcal{H}\}, \qquad \mathcal{H} + \mathcal{H}' = \{h + h' : h \in \mathcal{H},\ h' \in \mathcal{H}'\}.$$

Prove each statement first for a fixed sample, and then deduce the displayed expected-complexity statement by averaging over the sample.

(a) Show that $\mathcal{R}_m(\alpha\mathcal{H}) = |\alpha|\mathcal{R}_m(\mathcal{H})$.

(b) Show that $\mathcal{R}_m(\mathcal{H} + \mathcal{H}') = \mathcal{R}_m(\mathcal{H}) + \mathcal{R}_m(\mathcal{H}')$.

(c) Let $\mathcal{M} = \{x \mapsto \max\{h(x), h'(x)\} : h \in \mathcal{H},\ h' \in \mathcal{H}'\}$. Show that

$$\mathcal{R}_m(\mathcal{M}) \leq \mathcal{R}_m(\mathcal{H}) + \mathcal{R}_m(\mathcal{H}').$$

Hint: use $\max\{a,b\} = \frac{1}{2}(a + b + |a - b|)$. You may use the following contraction inequality without proof: for an L_ϕ-Lipschitz function ϕ with $\phi(0) = 0$,

$$\widehat{\mathcal{R}}_S(\phi \circ \mathcal{A}) \leq L_\phi \widehat{\mathcal{R}}_S(\mathcal{A}), \qquad \phi \circ \mathcal{A} = \{\phi \circ a : a \in \mathcal{A}\}.$$

Problem 4. Regularized neural networks (FML Exercise 3.11)

Use the convention in Problem 3. Fix $n_1, n_2, m \geq 1$ and a sample $S = (x_1, \ldots, x_m) \subseteq \mathbb{R}^{n_1}$. Let $\Lambda_0, \Lambda \geq 0$ and let $\sigma : \mathbb{R} \to \mathbb{R}$ be L-Lipschitz with $\sigma(0) = 0$, as for ReLU or tanh. Consider

$$\mathcal{H} = \left\{ x \mapsto \sum_{j=1}^{n_2} w_j \sigma(u_j^\top x) : \|w\|_1 \leq \Lambda_0, \quad \|u_j\|_2 \leq \Lambda \text{ for every } j \right\}.$$

The networks in this problem have no bias parameters. Define

$$\mathcal{H}_{\text{lin}} = \{x \mapsto s\, u^\top x : \|u\|_2 \leq \Lambda,\ s \in \{-1, +1\}\}.$$

(a) Show that

$$\widehat{\mathcal{R}}_S(\mathcal{H}) = \frac{\Lambda_0}{m} \mathbb{E}_\epsilon \sup_{\|u\|_2 \leq \Lambda} \left| \sum_{i=1}^m \epsilon_i \sigma(u^\top x_i) \right|.$$

Hint: optimize over the output weights w first.

(b) You may use the following absolute-value contraction inequality without proof: if $\phi(0) = 0$ and ϕ is L_ϕ-Lipschitz, then

$$\mathbb{E}_\epsilon \sup_{g \in \mathcal{A}} \left| \frac{1}{m} \sum_i \epsilon_i \phi(g(x_i)) \right| \leq 2 L_\phi \mathbb{E}_\epsilon \sup_{g \in \mathcal{A}} \left| \frac{1}{m} \sum_i \epsilon_i g(x_i) \right|.$$

Deduce that $\widehat{\mathcal{R}}_S(\mathcal{H}) \leq 2L\Lambda_0 \widehat{\mathcal{R}}_S(\mathcal{H}_{\text{lin}})$. Explain where symmetry of the linear class is used.

(c) Prove that

$$\widehat{\mathcal{R}}_S(\mathcal{H}_{\text{lin}}) = \frac{\Lambda}{m} \mathbb{E}_\epsilon \left\| \sum_{i=1}^m \epsilon_i x_i \right\|_2.$$

(d) Use Cauchy–Schwarz or Jensen's inequality and independence of the signs to show that

$$\widehat{\mathcal{R}}_S(\mathcal{H}_{\text{lin}}) \leq \frac{\Lambda}{m} \sqrt{\sum_{i=1}^m \|x_i\|_2^2}.$$

(e) If $\|x_i\|_2 \leq r$ for every i, derive an upper bound for $\widehat{\mathcal{R}}_S(\mathcal{H})$ in terms of $L, \Lambda_0, \Lambda, r, m$. State whether this bound depends on the hidden-layer width n_2.