MATH/COMP 562 FINAL EXAM STUDY GUIDE
VERSION: April 13, 2023

ADAM M. OBERMAN

Instructions. This is a study guide for the final exam. Final exam questions will be based on the following material.

5.1. **PAC Learning bounds: Theory and Examples.**

Exercise 5.1 (PAC Learning). Mohri Example 2.4 [axis-aligned rectangle] or Mohri Exercise 2.2. [higher dimensional rectangles]

Exercise 5.2. Know definition of VC dimension, and be able to apply it on examples such as

- VC dimension of axis-aligned rectangles (4).
- VC dimension of axis-aligned squares = 3
- VC dimension of all squared in $\mathbb{R}^2 = 5$
- VC dimension of half spaces in $\mathbb{R}^d$ is at least $d+1$

5.2. **Concentration of Measure.**

Exercise 5.3 (Apply McDiarmid's inequality in Rademacher Complexity proof). $\Phi(S) = \sup_{h\in\mathcal{H}} \Big(\mathbb{E}[h] - \widehat{\mathbb{E}}_S[h]\Big)$. Show $\Phi(S) - \Phi(S') \leq 1/m$, then Apply McDiarmid's inequality to Φ as in Mohri (3.7)

5.3. **Rademacher Complexity.**

- Study the proof of Theorem 3.3 Mohri

Exercise 5.4 (HW 2, exercise 1.2, Rademacher complexity as upper bound on generalization gap). Given a hypothesis class $\mathcal{H}$, of functions $h : X \to \mathbb{R}$, and a dataset $S = \{x_1, \dots, x_m\} \subset X$. Define $\Phi(S) = \sup_{h\in\mathcal{H}} \Big(\mathbb{E}[h] - \widehat{\mathbb{E}}_S[h]\Big)$ to be the least upper bound for the generalization gap of a function in $\mathcal{H}$. Prove that $\mathbb{E}_S[\Phi(S)] \leq 2\mathfrak{R}_m(\mathcal{H})$

Exercise 5.5 (HW 2, exercise 1.1, Rademacher complexity of linear hypothesis). Prove the following theorem.

Theorem 5.1 (Theorem 5.10 of Mohri). *Define $B_r = \{x \in \mathbb{R}^d \mid \|x\| \le r\}$. Let $S = \{x_1, \dots x_m\} \subset X \subset \mathbb{R}^d$ with $X = B_r$. Consider the linear hypotheses $h(w, x) = w \cdot x$ and set $\mathcal{H} = \{h(x, w) = w \cdot x \mid x \in X, w \in B_\lambda\}$. Prove that the empirical Rademacher complexity is bounded as follows,*

$$\widehat{\mathfrak{R}}_S(\mathcal{H}) \le \frac{r\lambda}{\sqrt{m}}$$

5.4. Convex Learning Problems.

- Define convex / smooth / Lipschitz functions
- Show that a given function is convex / smooth / Lipschitz
- Show that a standard ML problem (logistic regression, classification with margin loss / regression) is convex / Lipschitz / smooth

Exercise 5.6 (Convex Learning problems). Shalev Schwartz 12.2 / Shalev Schwartz 12.3.

5.5. Stable Learning Problems: proof of stability.

Exercise 5.7. SS Corollary 13.6

Exercise 5.8 (Study HW 3.7.: Algorithmic Stability for Ridge Regression). Some subset of 3.7 could be an exam problem

5.6. Stable Learning Problems: Concentration of Measure.

Exercise 5.9 (First part of Proof of Theorem 14.2 Mohri). Define $\Phi(S) = R(h_S) - \widehat{R}_S(h_S)$. Show Φ satisfied the bounded differences inequality, and apply McDiarmid's inequality. to obtain (14.2 / 14.3).

Exercise 5.10 (Second part of Proof of Theorem 14.2 Mohri). Define $\Phi(S) = R(h_S) - \widehat{R}_S(h_S)$ Prove $E_S[\Phi(S)] \le \beta$

5.7. Generative Modeling.

- Not covered on the final exam

5.8. Contraction mappings in the maximum norm.

- The MDP and Bellman equation will not be covered on the final exam
- Contraction mapping theorem (from course notes) will be covered. Example questions below.

Exercise 5.11. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a strict γ-contraction. Given x^0, define $x^{n+1} = T(x^n)$. Let x^* be the unique fixed point of T. Prove that $\|x^n - x^*\| \le \gamma^n \|x^0 - x^*\|$.

Exercise 5.12. Let $w_1, w_2 \in \mathbb{R}^d$. Suppose $w_1, w_2 \ge 0$ and $\sum_{i=1}^d (w_1)_i = \sum_{i=1}^d (w_2)_i = \gamma < 1$.

(a) Prove that $\gamma \min_i x_i \le w_1 \cdot x \le \gamma \max_i x_i$

(b) Let $r_1, r_2 \in R$. Prove that $\max(w_1 x + r_1, w_2 x + r_2)$ is a contraction in the max norm.

(c) Let W be a $d \times d$, matrix, with nonnegative coefficients. Let $\gamma < 1$. Suppose that $\sum_{i=1}^{d} w_{ji} \leq \gamma$ for each $j = 1, \ldots, d$. Let $r \in \mathbb{R}^d$. Define $g(x) = Wx + r$. Prove that g is a γ contraction, and conclude that g has a unique fixed point.