MATH/COMP 562 ASSIGNMENT 3
DUE: MARCH 26TH (SUNDAY)
VERSION: March 21, 2023

ADAM M. OBERMAN

Instructions. Submit your solutions on MyCourses course page. You can get help from other students, but you should do the write up yourself. Code solutions can be a PDF of a workbook, you can use any programming language, but NumPy is recommended.

VC dimension. Refer to "Foundations of Machine Learning," by Mohri, Chapter 3.

Exercise 3.1 (Rademacher identities). Mohri Exercise 3.8. part (c) only

Exercise 3.2 (VC dimension of finite hypothesis sets). Mohri Exercise 3.14.

Exercise 3.3 (VC dimension of closed balls). Mohri Exercise 3.17.

Exercise 3.4 (function class with infinite VC dimension). Mohri Exercise 3.20.

Stability Theory. Refer to posted class notes and Ch 13 of Understanding Machine Learning (Shalev-Shwartz).

Exercise 3.5 (Strong Convexity). Recall that a function f is λ-strongly convex if

$$\alpha f(w) + (1-\alpha) f(u) \geq f(\alpha w + (1-\alpha)u) + \frac{\lambda}{2}\alpha(1-\alpha)\|w-u\|^2, \quad \forall u, w \in \mathbb{R}^d, \alpha \in [0,1]$$

(a) Prove that $f(w) = \|w\|^2$ is λ-strongly convex with $\lambda = 2$.
(b) Give an example of a convex function which is not λ-strongly convex for any $\lambda > 0$.
(c) Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex. Prove that $g(w, \lambda) = f(w) + \lambda\|w\|^2/2$ is λ-strongly convex.
(d) Prove that g is λ-strongly convex if and only if $f(w) = g(w) - \lambda\|w\|^2/2$ is convex.

Exercise 3.6 (Random Variables and Strong Convexity). Let $X \in \mathbb{R}$ be a discrete ~~vector-valued~~ random variable, with distribution

$$\rho(X) = \begin{cases} w, & \text{with probability } \alpha \\ u, & \text{with probability } (1-\alpha) \end{cases}$$

(1) Find $\mathbb{E}[X]^2, \mathbb{E}[X^2], Var(X)$. Show that the expression $\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = Var(X)$ corresponds to
$$\alpha f(w) + (1-\alpha) f(u) - f(\alpha w + (1-\alpha)u) = \alpha(1-\alpha)\|w-u\|^2$$
with $f(w) = w^2$.
(2) Suppose f is λ strongly convex. Let X be as above. Show that
$$\mathbb{E}[f(X)] - f(\mathbb{E}[X]) \geq \frac{\lambda}{2} Var(X)$$

Exercise 3.7 (Algorithmic Stability for Ridge Regression updated for clarity). Write,

$$f(w, S) = \widehat{L}_S(w) = \frac{1}{m}\sum_{i=1}^{m}(w \cdot x_i - y_i)^2$$

Consider the datasets S_1, S_2, which differ by exactly one data point. Write
$$z^1 = (x^1, y^1) \in S_1, \quad z^2 = (x^2, y^2) \in S_2$$
for the different points. Define $g(w, S) = f(w, S) + \lambda\|w\|^2/2$ and
$$w_i = A(S_i) = \arg\min_w g(w, S_i), \quad i = 1, 2$$

(a) Explain why $g(w, S)$ is λ-strongly convex (as a function of w).
(b) Explain how strong convexity is applied to obtain

$$g(w_2, S_1) - g(w_1, S_1) \geq \frac{\lambda}{2}\|w_1 - w_2\|^2$$
$$g(w_1, S_2) - g(w_2, S_2) \geq \frac{\lambda}{2}\|w_1 - w_2\|^2$$

(c) Simplify to obtain

$$\lambda\|w_1 - w_2\|^2 \leq \frac{1}{m}\left(\ell(w_2, z^1) - \ell(w_1, z^1) + \ell(w_1, z^2) - \ell(w_2, z^2)\right)$$

for $\ell(w, z) = (w \cdot x - y)^2$

(d) Assume that $\|x\| \leq C_x$ and $|y| \leq C_y$ for all data x, y, and that $\|w_1\|, \|w_2\| \leq C_w$. Bound the RHS of the previous inequality. (Hint: apply the identity $(a^2 - b^2) = (a - b)(a + b)$ to the loss.)

(e) Conclude that the Ridge Regression problem is replace one stable (ROS) and give the rate as a function of λ, m.

Coding exercises.

Exercise 3.8. Using your choice of dataset (one dimensional is fine), perform Ridge regression. Find $w = A(S)$. Change the one data point to obtain S'. Compute $w' = A(S')$. Changing the data by different amounts, check the stability bounds. Plot the $\|w - w'\|$ against the bound.