MATH 462 ASSIGNMENT 5
VERSION November 29, 2023

ADAM M. OBERMAN

Instructions. You are also free to consult the textbook references or the internet, but be careful of conflicting notation and definitions. Submit your work on the MyCourses course page.

Math Solutions are preferred handwritten, but typed (good quality latex) is permitted. 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 language.

5.1. **Convex Learning Problems.** Refer to Ch 12 of Understanding Machine Learning (Shalev-Shwartz).

For the next two exercises, consider the classification problem with $x \in X = [-1,1]$, $y \in \{\pm 1\}$. Let m be the size of S, which is given by

$$S = \{(-1,-1), (-0.8,-1), (-0.6,+1), (-0.3,-1), (0.3,+1), (0.6,-1), (0.8,+1), (+1,+1)\}$$

Let $h(x,w) = x - w$, and $c(h) = \operatorname{sgn}(h)$.

Exercise 5.1 (non-convexity of 0-1 loss). With S and h as above, consider the zero-one loss

$$L(w) = \frac{1}{m}\sum_{i=1}^{m} \ell_{0-1}(c(x_i - w), y_i)$$

(a) Plot (sketch by hand) the function $L(w)$, for $w \in [-1,1]$.
(b) Identify the local minima (they can be intervals), and the global minimum value of the loss function.

Exercise 5.2 (Convex classification loss). With S and h as above, consider the classification loss $\ell_{\log}(h,y) = \log(1+\exp(-yh))$, along with the loss function

$$L(w) = \frac{1}{m}\sum_{i=1}^{m} \ell_{\log}(h(x_i,w), y_i)$$

(a) Plot (or sketch by hand) the function $L(w)$, for $w \in [-1,1]$.
(b) For any values of (x,y) in the domain $X \times \{\pm 1\}$, consider the function $g(w) = \ell_{\log}(h(x,w),y)$. Find the first and second derivatives of g. Conclude that g is a convex function of w.
(c) Explain why $L(w)$ is also a convex function of w. Which properties of convex functions did you use?

Exercise 5.3 (Convexity, Lipschitz, and Smoothness of logistic regression loss.). Shalev-Shwartz Problem 12.2.

Exercise 5.4 (Lipschitz continuity of the hinge loss). Shalev-Shwartz Problem 12.3.

5.2. **Feature Regression and orthogonal features.**

Exercise 5.5. Consider the following vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

(a) Make these vectors orthogonal using the Gram-Schmidt process, by performing the following steps. (i) Start with the first vector and set $\mathbf{u_1} = \mathbf{v_1}$. (ii) Next, compute the projection of $\mathbf{v}_2$ onto $\mathbf{u}_1$ and subtract it from $\mathbf{v}_2$:

$$\mathbf{u}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2), \quad \text{where } \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2) = \frac{\mathbf{u}_1 \cdot \mathbf{v}_2}{\|\mathbf{u}_1\|^2} \cdot \mathbf{u}_1$$

(iii) Lastly, compute the projection of $\mathbf{v}_3$ onto the subspace spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$, and subtract it from $\mathbf{v}_3$:

$$\mathbf{u}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_3)$$

Find the orthogonalized vectors $\mathbf{u}_1, \mathbf{u}_2$, and $\mathbf{u}_3$.

(b) Next, given the vector $y = [6, 0 - 6]^\top$. Solve

$$\min_w \|Fw - y\|^2, \quad F = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3], \quad Fw = \sum_i w_i \mathbf{u}_i$$

with F the 3×3 matrix with columns given by the vectors $\mathbf{u}_i$, . Express w_i in terms of inner products of y with certain vectors.

Exercise 5.6. Now consider the vector space of functions $u : X = [0,1] \to \mathbb{R}$, with the inner product $(f,g) = (f,g)_X = \int_0^1 f(x)g(x)dx$. Start with the functions

$$v_1(x) = 1, \quad v_2(x) = x, \quad v_3(x) = x^3$$

(a) Make these functions orthogonal, using the Gram-Schmidt process for these functions, with same ordering as in the previous exercise, to find the orthogonalized functions u_1, u_2, u_3.
(b) Now, assuming u_1, u_2, u_3 are orthogonal, given a function y, consider the functional regression problem

$$\min_w \|h(x,w) - y(x)\|_X^2$$

where $h(x,w) = w_1u_1(x) - w_2u_2(x) + w_3u_3(x)$. Express the coefficients, w_1, w_2, w_3 of the minimizer, $h(x,w)$, in terms of inner products of y and the functions u_i. For $y(x) = \exp(5x)$, find w_1.

Exercise 5.7. Let $S = \{(x_1,y_1), \dots (x_m,y_m)\}$ where $X = [x_1,\dots,x_m]^\top = [0.01, 0.02, \dots, 0.99, 1.00]^\top$ is $m = 100$ equally spaced points in $[0,1]$. Let $y = \exp(5x)$ and let $y_i = y(x_i)$. Define the feature regression problem

$$\min_w L(w,h,S) = \frac{1}{m}\sum_{i=1}^m \|y_i - h(w,x_i)\|^2$$

(a) Solve the feature regression problem using features $f(x) = [1, x, x^3]$ and with $h_1(x,w) = w \cdot f(x)$.
(b) Set u_i to be the orthogonal features from the previous problem. Set $h_2(x,v) = v \cdot g(x)$ where $g(x) = [u_1(x), u_2(x), u_3(x)]$. Solve the feature regression problem with h_2. Are the solutions equal, $h_2(x,v) = h_1(x,w)$ as functions of x?
(c) Now let $h_3(x,w)$ be the solution of the function regression problem with $X = [0,1]$. Approximate it by taking $m_2 = 1,000$ and solving the regression problem on the larger dataset. Plot $e(x) = h_1(x) - h_3(x)$, and find the mean squared error,

$$E(h_1,h_3,X) = \frac{1}{m}\sum_{i=1}^m (h_1(x_i) - h_3(x_i))^2$$

What is the value of the mean squared error? How does it compare to m (e.g. $1/m$, $1/m^2$)?