

# MATH 462 ASSIGNMENT 1

## VERSION September 18, 2021

ADAM M. OBERMAN

### 1. EXERCISES

#### 1.1. Example calculations.

- (1) (One variable quadratic regression) Consider (EL) with  $X = [1, 2, 3]^T$ ,  $y = [2, 4, 5]^T$ , linear model  $h_w(x) = w * x$ , and quadratic loss, (EL1d). Solve the problem by finding a critical point,  $\hat{L}'(w) = 0$ .
- (2) (One variable  $\ell_1$  regression) Same setup as in the previous problem: consider (EL) with  $X = \{1, 2\}$   $Y = \{1, 3\}$ . Use linear model  $h = w * x$ . Solve the problem with the  $\ell_1$  loss. Hint: plot the function  $\hat{L}(w)$ , which is piecewise linear, and find the minimum value (by finding the intersection of two lines).
- (3) ( $\ell_1$  central value) Plot (by hand, or otherwise),  $\hat{L}(w) = \frac{1}{m} \sum_i |w - y_i|$  in the case  $y = [2, 4, 5]^T$ , and in the case  $y = [3, 6]$ . Find all the minimizers in both cases.
- (4) (Polynomial regression, coding). Suppose our data points are  $x = [0, .1, \dots, .9, 1]$ . Let  $f(x) = (1, x)$  (affine linear regression). Set  $y = \sin(2\pi x) + .1e$ , where  $e$  is uniformly random data on  $[0, 1]$ . (i) Set up the data matrix,  $F$ . What are the sizes of  $F$ ,  $F^T F$ , and  $w$ ? Plot the error and solution. Is the fit good? (ii) Redo the problem with  $f(x) = (1, x, x^2, x^3)$ .
- (5) (Two variable quadratic regression) Set  $X = \{(3, 0), (0, 2), (1, 1)\}$   $Y = \{6, 2, 5\}$  Setup the quadratic regression problem (i) by minimizing (EL) directly (i.e. take derivatives with respect to  $w_1$  and  $w_2$  and solve. (ii) by setting up the matrix equation (Lin) and solving it.

- (6) Consider [https://en.wikipedia.org/wiki/Winsorized\\_mean](https://en.wikipedia.org/wiki/Winsorized_mean). Give an example with 10 numbers where the 10% Winsorized mean is the same as the minimizer of the Huber loss (with, say  $\delta = 1$ ). Explain the main difference between the Winsorized mean and the minimizer of the Huber loss? (Hint: the Huber loss has a scale  $\delta$  which determines the outliers, but the Winsorized mean has a fraction of values).
- (7) (Compare the regression loss functions) Consider  $y_1, \dots, y_{10}$  consisting of nine 0 and one value  $y$ . What is the solution (as a function of  $y$ ) of (CVIL) for each of the three main regression losses (take  $\delta = 1$  in the Huber).

### 1.2. Theory exercises.

- (1) Consider (Lin) when  $d = 1$ . Solve the equation, for  $w$ , and give the solution using vector notation.
- (2) Give a different derivation of (Lin) by citing the matrix theory fact that  $\min_w \|Xw - y\|^2$  is given by (Lin).
- (3) (Huber loss) Show that the Huber loss is continuous, and differentiable. Find the second derivative of the function.
- (4) *Problem removed*
- (5) Verify that the three main regression losses are all convex.

### 1.3. Loss design exercises.

- (1) (Loss designs for grading scheme) Consider a grading scheme where there are five assignments. Suppose we want a grading scheme that is less sensitive to outliers, e.g. with a score of .9, .9, .9, 0 (one missed assignment) we don't want the hard penalty given by the average. At the same time, we want every grade to have a small effect (to encourage performance when possible). (i) propose a simple scheme to do this. (ii) Suppose we want a missed assignment to have an effect of no more than  $\delta = .2$  on the average. Show that solving (CVIL) using the Huber loss with  $\delta = .2$  accomplishes this (at least in the example above).
- (2) (Loss design: flipped huber) Design a 'flipped' Huber loss function, which is quadratic for  $|t| \geq \delta$  and equals  $|t|$  for otherwise. Set up the quadratic so that the loss is continuously differentiable.
- (3) *Problem removed*

## 2. REFERENCED FORMULAS

$$(EL) \quad \widehat{L}(w) = \frac{1}{m} \sum_{i=1}^m \ell(h_w(x_i), y_i)$$

$$(EL1d) \quad \widehat{L}(w) = \frac{1}{m} \sum_{i=1}^m q(wx_i - y_i).$$

$$(Lin) \quad X^T X w = X^T y$$

$$(CVIL) \quad \min_w \frac{1}{m} \sum_{i=1}^m \ell(w, y_i)$$