Lab 2: Empirical Risk and Simple Linear Regression

EECS 245, Spring 2026 at the University of Michigan

due for completion at 11:59PM Ann Arbor Time on Monday, May 11th, 2026

Name: _____

uniqname: _____

Each lab worksheet will contain several activities, some of which will involve writing code and others that will involve writing math on paper. To receive credit for a lab, you must complete as many of the activities as you can in 2 hours and submit a PDF of your work to Gradescope. We will provide specific instructions on how to submit programming activities (e.g. submitting the notebook or including a screenshot of some output).

Feel free to work with others in the course, but you must submit individually.

Recap: The Modeling Recipe

In Chapter 1.3, we introduced the three-step modeling recipe for finding optimal model parameters, which ultimately helps us make the best possible predictions.

1. **Choose a model.**

$$\underbrace{h(x_i) = w}_{\text{constant model}} \qquad \underbrace{h(x_i) = w_0 + w_1 x_i}_{\text{simple linear regression model}}$$

2. **Choose a loss function.**

$$\underbrace{L_{\text{sq}}(y_i, h(x_i)) = (y_i - h(x_i))^2}_{\text{squared loss}} \qquad \underbrace{L_{\text{abs}}(y_i, h(x_i)) = |y_i - h(x_i)|}_{\text{absolute loss}}$$

3. **Minimize *average loss* (also called *empirical risk*) to find optimal model parameters.**

- Constant model, squared loss: $R_{\text{sq}}(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w)^2 \implies w^* = \bar{y}$
- Constant model, absolute loss: $R_{\text{abs}}(w) = \frac{1}{n}\sum_{i=1}^{n}|y_i - w| \implies w^* = \text{Median}(y_1, y_2, \ldots, y_n)$
- Simple linear regression model, squared loss:

$$R_{\text{sq}}(w_0, w_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i))^2 \implies w_1^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad w_0^* = \bar{y} - w_1^* \bar{x}$$

Activity 1: Relative Squared Loss

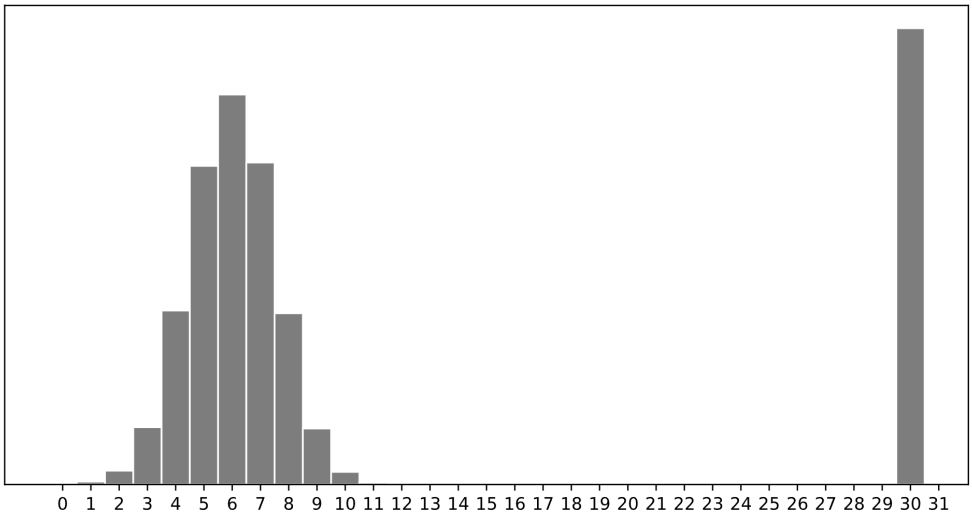
Suppose we'd like to find the optimal parameter, w^*, for the constant model $h(x_i) = w$. To do so, we use the following loss function, called the **relative squared loss**:

$$L_{\text{rsq}}(y_i, h(x_i)) = \frac{(y_i - h(x_i))^2}{y_i}$$

What value of w minimizes the average loss (i.e. empirical risk) when using the relative squared loss function – that is, what is w^*? Your answer should only be in terms of the variables $n, y_1, y_2, \ldots, y_n$, and any constants.

Activity 2: Rapid Fire

Consider a dataset of n **integers**, $y_1, y_2, \ldots, y_n$, whose histogram is given below:



a) Which of the following is closest to the constant prediction w^* that minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}\begin{cases}0 & y_i = w\\ 1 & y_i \neq w\end{cases}$$

◯ 1 ◯ 5 ◯ 6 ◯ 7 ◯ 11 ◯ 15 ◯ 30

b) Which of the following is closest to the constant prediction w^* that minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - w|$$

◯ 1 ◯ 5 ◯ 6 ◯ 7 ◯ 11 ◯ 15 ◯ 30

c) Which of the following is closest to the constant prediction w^* that minimizes:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - w)^2$$

◯ 1 ◯ 5 ◯ 6 ◯ 7 ◯ 11 ◯ 15 ◯ 30

d) Which of the following is closest to the constant prediction w^* that minimizes:

$$\lim_{p\to\infty}\frac{1}{n}\sum_{i=1}^{n}|y_i - w|^p$$

Hint: Think about the effect of outliers.

◯ 1 ◯ 5 ◯ 6 ◯ 7 ◯ 11 ◯ 15 ◯ 30

Activity 3: Slope of Mean Absolute Error

Consider a dataset of 8 points, $y_1, y_2, \ldots, y_8$ that are in sorted order, i.e. $y_1 < y_2 < \ldots < y_8$.

Recall that mean absolute error, $R_{\text{abs}}(w)$, for the constant model $h(x_i) = w$ is defined as:

$$R_{\text{abs}}(w) = \frac{1}{n} \sum_{i=1}^{n} |y_i - w|$$

This is a piecewise linear function that changes slope at each data point. The slope of $R_{\text{abs}}(w)$ at any w that is not a data point is:

$$\frac{\text{d}}{\text{d}w} R_{\text{abs}}(w) = \frac{\text{\# left of } w - \text{\# right of } w}{n}$$

Suppose that $y_4 = 10$, $y_5 = 14$, $y_6 = 22$, and $R_{\text{abs}}(11) = 9$. What is $R_{\text{abs}}(22)$?

Hint: You don't have all 8 of the y-values, so you can't find $R_{abs}(22)$ just by plugging in numbers into the formula for $R_{abs}(w)$. Instead, think about how to use the slope formula.

Activity 4: Programming

Complete the tasks in the `lab02.ipynb` notebook.

There are two ways to access the supplemental Jupyter Notebook:

- **Option 1 (preferred)**: Set up a Jupyter Notebook environment locally, use `git` to clone our **course repository**, and open `labs/lab02/lab02.ipynb`. For instructions on how to do this, see the **Environment Setup** page of the course website.
- **Option 2**: Click **here** to open `lab02.ipynb` on DataHub. Before doing so, read the instructions on the **Environment Setup** page on how to use the DataHub.

Once you're done, include a screenshot of your completed Activity 4 implementation in your PDF submission of Lab 2 to Gradescope, making sure to include proof that the (local) autograder passed.

Activity 5: Reverse Regression

Suppose we have a dataset of n houses that were recently sold in the Ann Arbor area. For each house, we have its square footage and most recent sale price. The correlation between square footage and price is r.

First, we minimize mean squared error to fit a simple linear model that uses square footage to predict price. The resulting regression line has an intercept of w_0^* and slope of w_1^*.

$$\text{predicted price}_i = w_0^* + w_1^* \cdot \text{square footage}_i$$

We're now interested in minimizing mean squared error to fit a simple linear model **that uses price to predict square footage** — that is, we're "reversing" the x and y variables. Suppose this new regression line has an intercept of β_0^* and slope of β_1^*.

Find β_1^*. Give your answer in terms of one or more of n, r, w_0^*, and w_1^*.

Activity 6: Transformed Data

Suppose we're given a dataset of n points, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where $\bar{x}$ is the mean of $x_1, x_2, \ldots, x_n$ and $\bar{y}$ is the mean of $y_1, y_2, \ldots, y_n$.

Using this dataset, we create a *transformed* dataset of n points, $(x_1', y_1'), (x_2', y_2'), \ldots, (x_n', y_n')$, where:

$$x_i' = 4x_i - 3 \qquad y_i' = y_i + 24$$

So the transformed dataset is of the form

$$(4x_1 - 3, y_1 + 24), (4x_2 - 3, y_2 + 24), \ldots, (4x_n - 3, y_n + 24)$$

We decide to fit a simple linear model $h(x_i') = w_0 + w_1 x_i'$ on the transformed dataset using squared loss. We find that $w_0^* = 7$ and $w_1^* = 2$, so $h^*(x_i') = 7 + 2x_i'$.

a) Suppose we were to fit a simple linear model through the original dataset, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, again using squared loss. What would the optimal slope on the original dataset be?

b) Recall, the model $h^*(x_i') = w_0 + w_1 x_i'$ was fit on the transformed dataset, $(x_1', y_1'), (x_2', y_2'), \ldots, (x_n', y_n')$. $h^*(x_i')$ happens to pass through the point $(\bar{x}, \bar{y})$. What is the value of $\bar{x}$? Give your answer as an integer with no variables. *Hint: What else does $h^*(x_i')$ pass through?*

The following are extra practice. Don't feel pressured to answer all of these problems in lab, but make sure to attempt them at some point.

Activity 7: Relative Squared Loss, Continued

Recall the formula for **relative squared loss** from Activity 1:

$$L_{\text{rsq}}(y_i, h(x_i)) = \frac{(y_i - h(x_i))^2}{y_i}$$

a) Let $C(y_1, y_2, ..., y_n)$ be your minimizer w^* from Activity 1. That is, for a particular dataset $y_1, y_2, ..., y_n$, $C(y_1, y_2, ..., y_n)$ is the value of w that minimizes empirical risk for relative squared loss on that dataset.

What is the value of $\lim_{y_4 \to \infty} C(1, 3, 5, y_4)$ in terms of $C(1, 3, 5)$? Your answer should involve the function C and/or one or more constants.

Hint: To notice the pattern, evaluate $C(1, 3, 5, 100)$*,* $C(1, 3, 5, 10000)$*, and* $C(1, 3, 5, 1000000)$*.*

b) What is the value of $\lim_{y_4 \to 0} C(1, 3, 5, y_4)$? Again, your answer should involve the function C and/or one or more constants.

c) Based on the results of the previous two parts, when is the prediction $C(y_1, y_2, ..., y_n)$ robust to outliers? When is it not robust to outliers?

Activity 8: The Meaning of Mean Squared Error

Suppose we'd like to predict the number of minutes a delivery will take, y, as a function of distance, x. To do so, we look to our dataset of n deliveries, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, and fit two simple linear models:

- $F(x_i) = a_0 + a_1 x_i$, where:
$$a_1 = r\frac{\sigma_y}{\sigma_x}, \qquad a_0 = \bar{y} - a_1\bar{x}$$
Here, r is the correlation coefficient between x and y, $\bar{x}$ and $\bar{y}$ are their respective means, and σ_x and σ_y are their respective standard deviations.
- $G(x_i) = b_0 + b_1 x_i$, where b_0 and b_1 are chosen such that $G(x_i) = b_0 + b_1 x_i$ minimizes **mean absolute error** on the dataset. Assume that no other line minimizes mean absolute error on the dataset, i.e. that the values of b_0 and b_1 are unique.

a) Fill in the ???:

$$\sum_{i=1}^{n}(y_i - F(x_i))^2 \quad \boxed{???} \quad \sum_{i=1}^{n}(y_i - G(x_i))^2$$

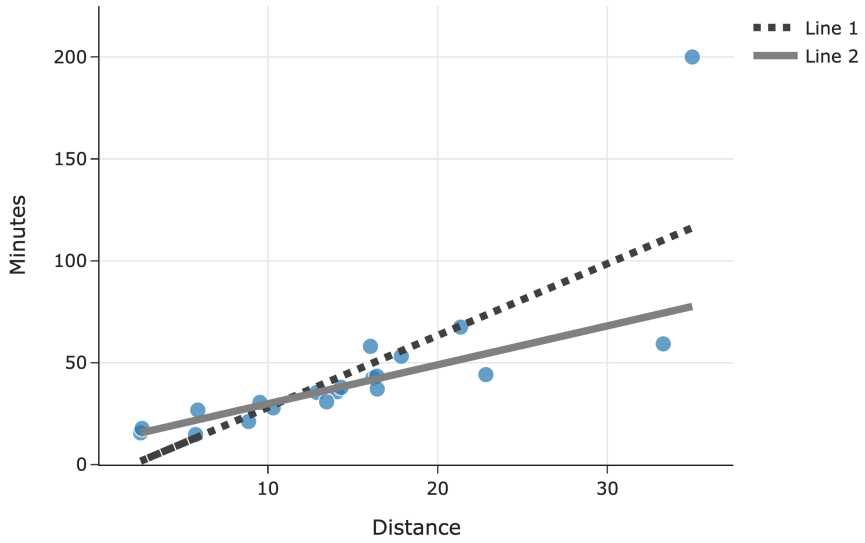
◯ $>$ ◯ $\geq$ ◯ $=$ ◯ $\leq$ ◯ $<$ ◯ Impossible to tell

b) Fill in the ???:

$$\left(\sum_{i=1}^{n}|y_i - F(x_i)|\right)^2 \quad \boxed{???} \quad \left(\sum_{i=1}^{n}|y_i - G(x_i)|\right)^2$$

◯ $>$ ◯ $\geq$ ◯ $=$ ◯ $\leq$ ◯ $<$ ◯ Impossible to tell

c) Below, we've drawn the lines for both F and G along with a scatter plot for the original n deliveries:



Which line corresponds to F? ◯ Line 1 ◯ Line 2

Activity 9: What Do You Mean?

Suppose we want to fit a simple linear model (using squared loss) that predicts the number of ingredients in a product given its price. We're given that:

- The average cost of a product in our dataset is \$40, i.e. $\bar{x} = 40$
- The average number of ingredients in a product in our dataset is 15, i.e. $\bar{y} = 15$

The intercept and slope of the regression line are $w_0^* = 11$ and $w_1^* = \frac{1}{10}$, respectively.

a) Suppose Victors' Veil (a skincare product) costs \$40 and has 11 ingredients. What is the squared loss of our model's predicted number of ingredients for Victors' Veil?

b) Is it possible to answer part **a)** above **just** by knowing $\bar{x}$ and $\bar{y}$, i.e. **without** knowing the values of w_0^* and w_1^*? Once you select an answer, explain it to your peers.

◯ Yes, it's possible ◯ No, it's not possible

Activity 10: A Refresher

Consider a dataset of $y_1, y_2, \ldots, y_n$, all of which are **positive**. We want to fit a constant model, $h(x_i) = w$, to the data.

Let w_p^* be the optimal constant prediction that minimizes average degree-p loss, $R_p(w)$, defined below:

$$R_p(w) = \frac{1}{n} \sum_{i=1}^{n} |y_i - w|^p$$

For example, w_2^* is the optimal constant prediction that minimizes $R_2(w) = \frac{1}{n} \sum_{i=1}^{n} |y_i - w|^2$

a) In each of the parts below, determine the value of the quantity provided. By "the data", we are referring to $y_1, y_2, \ldots, y_n$. The answer choices are as follows; **select one item in each row**.

A: The standard deviation of the data

B: The variance of the data

C: The mean of the data

D: The median of the data

E: The midrange of the data, $\frac{y_{\min} + y_{\max}}{2}$

F: The mode of the data

G: None of these

		A	B	C	D	E	F	G
(i)	h_0^*	◯	◯	◯	◯	◯	◯	◯
(ii)	h_1^*	◯	◯	◯	◯	◯	◯	◯
(iii)	$R_1(h_1^*)$	◯	◯	◯	◯	◯	◯	◯
(iv)	h_2^*	◯	◯	◯	◯	◯	◯	◯
(v)	$R_2(h_2^*)$	◯	◯	◯	◯	◯	◯	◯

b) Now, suppose we want to find the optimal constant prediction, h_{U}^*, using the "Ulta" loss function, defined below:

$$L_{\text{U}}(y_i, w) = y_i(y_i - w)^2$$

To find h_{U}^*, we minimize $R_{\text{U}}(w)$, the average Ulta loss. How does h_{U}^* compare to the mean of the data, M?

◯ $h_{\text{U}}^* > M$ ◯ $h_{\text{U}}^* \geq M$ ◯ $h_{\text{U}}^* = M$ ◯ $h_{\text{U}}^* \leq M$ ◯ $h_{\text{U}}^* < M$

c) Finally, to find the optimal constant prediction, we will instead minimize **regularized** average Ulta loss, $R_\lambda(w)$, defined below:

$$R_\lambda(w) = \left(\frac{1}{n} \sum_{i=1}^{n} y_i (y_i - w)^2 \right) + \lambda w^2$$

Here, assume $\lambda > 0$ is some positive constant. (We will cover regularization in more detail later in the term.)

Find w^*, the constant prediction that minimizes $R_\lambda(w)$. Give your answer as an expression in terms of the y_i's, n, and/or λ.