

# Problem Set 1\*

Riddhi Patel<sup>†</sup>

## Instruction

- To compile a  $\text{\LaTeX}$  file, you can use either the `pdflatex` command on your local machine or online tools like [Overleaf](#). If you're new to  $\text{\LaTeX}$ , make some effort to learn how to use it properly especially regarding math environments like [align](#).
- Create a new blank file (e.g., `hw1.tex`). Copy `hw-template.tex` into it. Don't change on this file directly, as you might want to reuse it.
- Delete this instruction part.
- Change the author info: name, email, netid.
- Change the homework info: name, due date, and students you worked with for this homework.
- For the coding part, you need to write your own code.
- When you finish your homework, download the pdf, and rename it before submitting to Canvas: `firstname-lastname-sol1.pdf`.
- The supplemented `header.tex` includes some useful packages and commands that make it easy to write things like  $\mathbf{E}_{x \sim p}[-\log p(x)]$ . Feel free to modify and define your own commands.
- *Do not* take photos of your handwritten equations and insert them as images. (Crazily enough, this has happened.) You must write equations entirely in  $\text{\LaTeX}$ . Any such photos will be counted as blank answers.

## Problem 1

**Answer:** Yes, it is possible to get  $\tilde{w}$  from  $w^*$  without retraining on the new data model.

Here is the mathematical proof as follows based on the information we have:

We know that our closed-form solution of  $w^*$  is:

$$w^* = (X^T X)^{-1} X^T y \quad (1)$$

---

\*Due: September 27, 2023, Student(s) worked with: Senuri Rupasinghe Collaborators

<sup>†</sup>NetID: rnp72 rnp72, Email: rnp72@scarletmail.rutgers.edu rnp72@scarletmail.rutgers.edu

We also know that we need to modify our y-label (known as our answer or value prediction) by:

$$y^{(i)} = ay^{(i)} + b \quad (2)$$

In order to prove that we can get  $\tilde{w}$  from  $w^*$ , we need to apply the change in the  $y^{(i)}$  to the formula above and substitute it into  $w^*$ 's closed-form solution and solve algebraically:

Note: b becomes B because it is a vector full of b constants.

$$\begin{aligned} \tilde{w} &= (X^T X)^{-1} X^T (ay^{(i)} + B) \\ &= (X^T X)^{-1} * X^T ay^{(i)} + X^T B \\ &= (X^T X)^{-1} X^T ay^{(i)} + (X^T X)^{-1} X^T B \\ &= a[(X^T X)^{-1} X^T y^{(i)}] + B[(X^T X)^{-1} X^T] \\ &= aw^* + B[(X^T X)^{-1} X^T] \end{aligned}$$

(3)

After substituting, distributing, and rearranging algebraically we are almost at our final answer. We now need to break down our B vector respectively and apply linear algebra properties to do so.

$$\begin{aligned} \tilde{w} &= aw^* + B[(X^T X)^{-1} X^T] \\ &= aw^* + bXe1[(X^T X)^{-1} X^T] \\ &= aw^* + bXe1[(X^T X)^{-1} X^T] \\ &= aw^* + be1(X^T X)^{-1} X^T X \\ &= aw^* + be1 \end{aligned}$$

(4)

We see that by breaking down vector B into 3 components,  $b^*X^*e1$  (where e1 represents the first column of matrix X where it's first dimension value is 1), while the rest we assume, be 0 ([1 . . .1]).

By rearranging into  $(X^T X)^{-1} * X^T X$  with linear algebra property, we will get an identity matrix and therefore will be canceled out of the equation.

$$\text{Final Answer: } g(w^*, a, b) = aw^* + br1$$

## Problem 2

**Answer:** Yes, it is possible to get  $\bar{w}$  from  $w^*$ .

This can be proven mathematically, shown below if we try to get  $\bar{w}$  from  $w^*$  using  $w^*$ 's closed form solution. We will go through the same algebraic steps from the previous problem, but this time substitute our  $X$  input matrices with  $xjcj$ :

$$\begin{aligned}\bar{w} &= (X^T X)^{-1} X^T y \\ &= [(cjxj^{(i)})^T * (cjxj^{(i)})]^{-1} * [(cjxj^{(i)})^T * y] \\ &= [(xj^{(i)T} cj^T) * (cjxj^{(i)})]^{-1} * [xj^{(i)T} (cj^T) * y] \\ &= (cjxj^{(i)})^{-1} * (xj^{(i)T} cj^T)^{-1} * (xj^{(i)T} cj^T * y) \\ &= xj^{(i)-1} cj^{-1} * cj^{-T} xj^{(i)-T} * xj^{(i)T} cj^T * y \\ &= cj^{-1} * [(xj^{(i)-T} * xj^{(i)-1}) * xj^{(i)T} * y] * (cj^{-T} * cj^T) \\ \bar{w} &= h(w^*, cd) = cj^{-1} * w^*\end{aligned}$$

(5)

By rearranging our values,  $(cj^{-T} cj^T)$  leads to an identity matrix according to linear algebra properties, so we are able to cancel it out of the equation. Furthermore, we were also able to rearrange our  $xj$  matrices with our  $y$  vector to reproduce  $w^*$  equation. As a result we manage to derive an equation of  $c^{-1} * w^*$ .

As long as  $w^*$  is present and our matrices are invertible (or full-rank) we should be able to retrieve  $\bar{w}$  without training on the new data model, as shown mathematically.

## Problem 3

**Answer:** Yes there is a closed-form solution possible through deriving the MLE equation as follows:

We have to replace  $\sigma$  with  $\sigma_i$  to represent our un-identical distribution with noise.

$$P(y|x; w, \sigma_i^2) = \arg \max_w \prod_{i=1}^N \frac{1}{\sigma_i^2 \sqrt{2\pi}}$$

$$\log(P(y, x; w, \sigma_i^2)) = \log\left(\frac{1}{\sigma_i^2 \sqrt{2\pi}} \exp\left(-\frac{(y-f(x;w))^2}{2\sigma_i^2}\right)\right)$$

$$\log(P(y, x; w, \sigma_i^2)) = -\log(\sigma_i^2 \sqrt{2\pi}) - \frac{1}{2\sigma_i^2}(y - f(x; w))^2$$

$$\hat{w}ML = \arg \max_w \sum_{i=1}^N \log p(y|x; w, \sigma_i^2) \quad (7)$$

$$= \arg \min_w \sum_{i=1}^N (y - f(x; w))^2 \quad (8)$$

Which then becomes

$$w^* = \arg \min_w \sum_{i=1}^N (y - f(x; w))^2 \quad (9)$$

The argmin simplifies to our closed form solution:

$$(X^T X)^{-1} X^T y$$

(10)