Redemption Homework

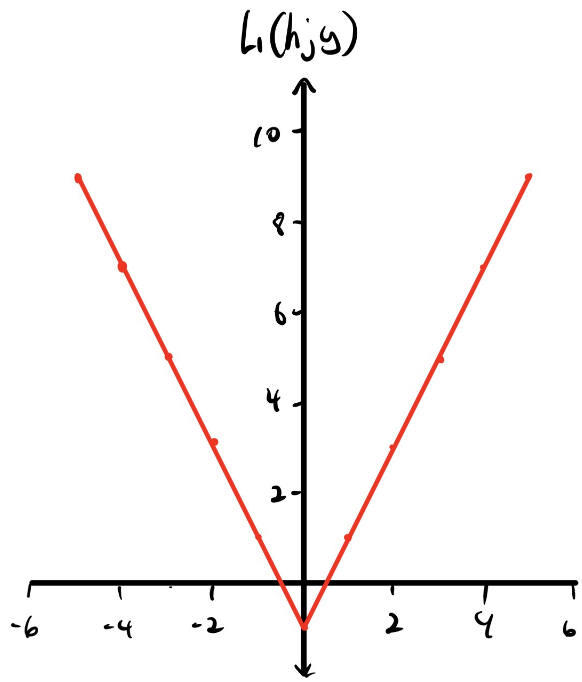
Jack Kai Lim

June 5, 2022

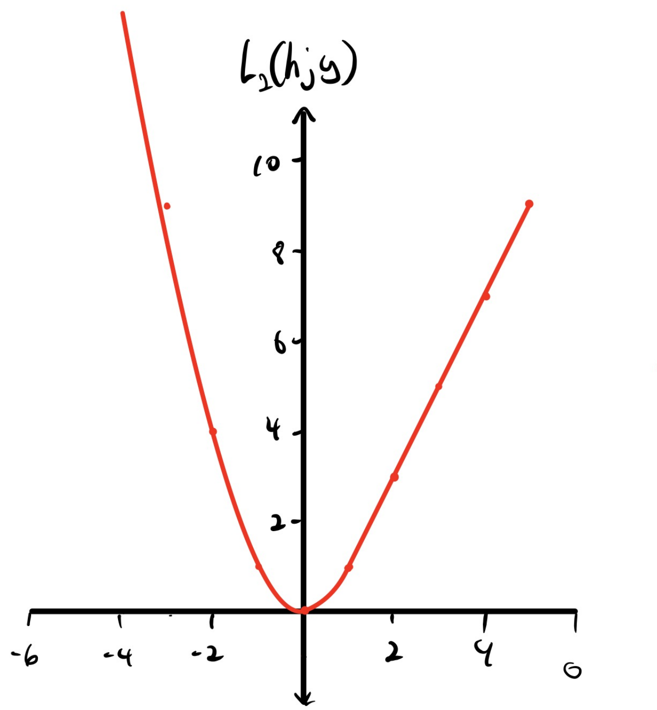
Problem 1

a)

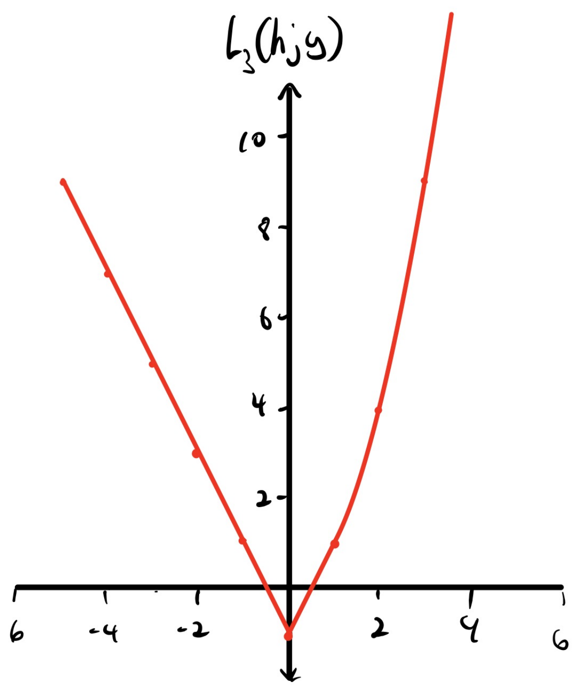
Loss Function 1, y = 0



Loss Function 2m y = 0



Loss Function 3, y = 0



b)

First we find the derivatives of the functions,

$$\frac{d}{dh}2|h-y|-1 = \frac{2(h-y)}{|h-y|}$$
$$\frac{d}{dh}((h-y)^2) = 2(h-y)$$

Using gradinet descent for loss 1 I got,

```python
def grad_descent1(h, data):
    n = np.size(data)
    grad = 0.0
    for i in range(n):
        grad = grad + (2*(h - data[i]))/abs(h - data[i])
    grad = grad/n
    return grad
```
[7] ✓ 0.5s Python

```python
#For Loss Function 1
h = 100
while abs(grad_descent1(h, y)) > 0.00001:
    h = h - grad_descent1(h, y)

h
```
[8] ✓ 0.6s Python

```
3.5999999999999988
```

For loss 2 I got,

```python
def grad_descent2(h, data):
    n = np.size(data)
    grad = 0.0
    for i in range(n):
        if h - data[i] > 1:
            grad = grad + (2*(h - data[i]))/abs(h - data[i])
        else:
            grad = grad + 2*(h - data[i])
    grad = grad/n
    return grad
```
[12] ✓ 0.3s Python

```python
#For Loss function 2
h = 100
while abs(grad_descent2(h, y)) > 0.00001:
    h = h - grad_descent2(h, y)

h
```
[13] ✓ 0.3s Python

```
5.153324896758235
```

And for loss 3 and the true value of theta is,

```python
def grad_descent3(h, data):
    n = np.size(data)
    grad = 0.0
    for i in range(n):
        if h - data[i] < 1:
            grad = grad + (2*(h - data[i]))/abs(h - data[i])
        else:
            grad = grad + 2*(h - data[i])
    grad = grad/n
    return grad
```

[14] ✓ 0.7s Python

```python
#For loss Function 3
h = 100
while abs(grad_descent3(h, y)) > 0.00001:
    h = h - grad_descent3(h, y)

h
```

[15] ✓ 0.4s Python

```
1.194002235097068
```

```python
theta
```

[16] ✓ 0.4s Python

```
2.888060372680481
```

First we note that θ is approximately equal to both the mean and median of the synthetic data set. And we know that $R_1(h)$ minimized gives us the median of a dataset therefore we get that $h_1 \approx \theta$ For $R_2(h)$ looking at the graph, we can note that all the values points on the left of y is on average larger than the values on the right. This causes the data to shift towards the right which makes $h_2 > \theta$. As for $R_3(h)$ using aa similar reasoning as $R_2(h)$ we can see that the values on the right are on average larger than the values on the left therefore this causes a shift towards the left causing $h_3 < \theta$

Problem 2

First we have that,

$$\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + ... + w_d^2} = (w_1^2 + w_2^2 + ... + w_d^2)^{\frac{1}{2}}$$

Now we compute the partial derivative with respect to w_i,

$$\begin{aligned}
\frac{d}{dw_i}\|\vec{w}\| &= \frac{1}{2}(w_1^2 + w_2^2 + ... + w_d^2)^{-\frac{1}{2}} \times 2w_i \\
&= \frac{1}{\sqrt{w_1^2 + w_2^2 + ... + w_d^2}} w_i \\
&= \frac{w_i}{\|\vec{w}\|}
\end{aligned}$$

Then we combine it for all the the components we get,

$$\begin{aligned}
\frac{d}{d\vec{w}}\|\vec{w}\| &= \begin{bmatrix} \frac{d}{dw_1}\|\vec{w}\| \\ \vdots \\ \frac{d}{dw_d}\|\vec{w}\| \end{bmatrix} \\
&= \begin{bmatrix} \frac{w_1}{\|\vec{w}\|} \\ \vdots \\ \frac{w_d}{\|\vec{w}\|} \end{bmatrix} \\
&= \frac{\vec{w}}{\|\vec{w}\|}
\end{aligned}$$

Problem 3

To prove the relationship firts we compute equation (1) using multivariate linear regression, using the design matrix

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_1^{(2)} \\ \vdots & \vdots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} \end{bmatrix}$$

We can get the normal equation of the linear regression using $X^T X\vec{w} = X^T\vec{y}$ which gets us,

$$\begin{bmatrix} 1 & \sum_{i=1}^n x_i^{(1)} & \sum_{i=1}^n x_i^{(2)} \\ \sum_{i=1}^n x_i^{(1)} & \sum_{i=1}^n {x_i^{(1)}}^2 & \sum_{i=1}^n x_i^{(1)} x_i^{(2)} \\ \sum_{i=1}^n x_i^{(2)} & \sum_{i=1}^n x_i^{(1)} x_i^{(2)} & \sum_{i=1}^n {x_i^{(2)}}^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i^{(1)} y_i \\ \sum_{i=1}^n x_i^{(2)} y_i \end{bmatrix}$$

As the first line of the system of linear equations ets us calculate w_0 of the linear regression we can use that to get 2 equations to prove the relationship.

$$w_0 + w_1\bar{x}^{(1)} + w_2\bar{x}^{(2)} = \bar{y}$$

Next using equation (2) we can get a relationship of w_0 in $w_0^{(1)}, w_2, \bar{x}^{(2)}$ which is,

$$\begin{aligned} w_0 + w_1\bar{x}^{(1)} + w_2\bar{x}^{(2)} &= w_0^{(1)} + w_1\bar{x}^{(1)} \\ w_0 &= w_0^{(1)} - w_2\bar{x}^{(2)} \end{aligned}$$

Which we can now use to prove the relationship by showing the equality on the second line of the system of linear equaton,

$$\begin{aligned} w_0 \sum_{i=1}^n x_i^{(1)} + w_1 \sum_{i=1}^n {x_i^{(1)}}^2 + w_2 \sum_{i=1}^n x_i^{(1)} x_i^{(2)} &= \sum_{i=1}^n x_i^{(1)} y_i \\ (w_0^{(1)} - w_2\bar{x}^{(2)})(n\bar{x}^(1)) + w_1 n{\bar{x}^{(1)}}^2 + w_2 n\bar{x}^{(1)}\bar{x}^{(2)} &= n\bar{x}^{(1)}\bar{y} \\ w_0^{(1)}(n\bar{x}^{(1)}) &= n\bar{x}^{(1)}\bar{y} - w_1 n\bar{x}^{(1)} \\ w_0^{(1)} &= \bar{y} - w_1\bar{x}^{(1)} \end{aligned}$$

The final line is equal to the given equation (2). We can perform the same steps to prove the thrid equation using the given equation (3),

$$\begin{aligned} w_0 + w_1\bar{x}^{(1)} + w_2\bar{x}^{(2)} &= w_0^{(2)} + w_2\bar{x}^{(2)} \\ w_0 &= w_0^{(2)} - w_1\bar{x}^{(1)} \end{aligned}$$

Using that we can prove the relationship on the third line of the linear equation,

$$
\begin{aligned}
w_0 \sum_{i=1}^{n} x_i^{(2)} + w_1 \sum_{i=1}^{n} x_i^{(1)} x_i^{(2)} + w_2 \sum_{i=1}^{n} {x_i^{(2)}}^2 &= \sum_{i=1}^{n} x_i^{(2)} y_i \\
(w_0^{(2)} - w_1 \bar{x}^{(1)})(n\bar{x}^{(2)}) + w_1 n \bar{x}^{(1)} \bar{x^{(2)}} + w_2 n {\bar{x}^{(2)}}^2 &= n\bar{x}^{(2)} \bar{y} \\
w_0^{(2)} n \bar{x}^{(2)} &= n\bar{x}^{(2)} \bar{y} - w_2 n {\bar{x}^{(2)}}^2 \\
w_0^{(2)} = \bar{y} - w_2 \bar{x}^2
\end{aligned}
$$

Which is equal to the given equation (3) therefore proving the relationship for the thrid linear system of equations. And therefore proving that the relationship is true as a whole.

b)

When the conditions given, using probability theory this tells us that there is no relationship between $x^{(1)}$ and $x^{(2)}$ and they are therefore independant.

Problem 4

When 2 events are independant, $P(A \cap B) = P(A)P(B)$ so we need to show that $P(A^c \cap B) = P(A^c)P(B)$ so we have,

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) \\ &= P(B)(1 - P(A)) \\ &= P(B)P(A^c) \end{aligned}$$

The first equality is reached as looking at a venn diagram is it easy to see that the area of $P(A^c \cap B)$ is the area of $P(B) - P(A \cap B)$. So as we have proven that $P(A^c \cap B) = P(A^c)P(B)$ using the same method, we can show $P(A \cap B^c) = P(A)P(B^c)$ is also true. Implying that $P(A^c)$ and $P(B^c)$ are independant of each other.

Problem 5

a)

As the states are all equally likely, the probability that the person is in denial is $\frac{1}{3}$ so we can calculate the probability by,

$$P(\text{Denial}|\text{Congratulated}) = \frac{\frac{5}{300}}{\frac{1}{3}(\frac{105}{100})} = \frac{1}{21}$$

Respectively we can calculate the probability for the other states,

$$P(\text{Fury}|\text{Congratulated}) = \frac{\frac{10}{300}}{\frac{1}{3}(\frac{105}{100})} = \frac{2}{21}$$

And,

$$P(\text{Acceptance}|\text{Congratulated}) = \frac{\frac{90}{300}}{\frac{1}{3}(\frac{105}{100})} = \frac{6}{7}$$

b)

Since we have different probabilities for the state of mind first we calculate,

$$P(\text{Congratulate}) = \frac{5}{100} \times \frac{80}{100} + \frac{10}{100} \times \frac{15}{100} + \frac{90}{100} \times \frac{5}{100} = \frac{1}{10}$$

Using that we can now calculate the probabilities,

$$P(\text{Denial}|\text{Congratulated}) = \frac{\frac{80}{100} \times \frac{5}{100}}{\frac{1}{10}} = \frac{2}{5}$$

$$P(\text{Fury}|\text{Congratulated}) = \frac{\frac{15}{100} \times \frac{10}{100}}{\frac{1}{10}} = \frac{3}{20}$$

$$P(\text{Acceptance}|\text{Congratulated}) = \frac{\frac{5}{100} \times \frac{90}{100}}{\frac{1}{10}} = \frac{9}{20}$$