Q1. $$\hat{\sigma}^2 = \frac{\sum (x_i - \hat{\mu})^2}{n}$$

(A.) Also, $\hat{\sigma}_j^2 = n^{-1} \sum_i x_{i,j}^2 - \bar{x}_j^2$

We have $s(D) = [S_1 \quad S_2 \quad S_3]^T$ where $\underset{3\times 1}{s(D)} = \begin{bmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 \\ n \end{bmatrix}$

$$D = \{x_1, x_2, \ldots\ldots, x_n\}$$

From the above definition of $s(D)$, we could simply have,

$$\frac{S_1}{S_3} = \frac{\sum_{i=1}^n x_i}{n}$$, which is nothing but the mean $\hat{\mu}$.

Now, let us consider $\frac{S_2}{S_3} - \left(\frac{S_1}{S_3}\right)^2$

$$\ldots S_3 \quad \left(S_3\right)$$

Putting in the values from $s(D)$ above,

$$\frac{S_2}{S_3} - \left(\frac{S_1}{S_3}\right)^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$$

which we could easily rewrite as

$$= E\left((X - \hat{u})^2\right) = \hat{\sigma}^2 \ (\underline{\text{variance}})$$

Note that $\hat{u} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

(B.)

<u>Given</u> :- $s^2 = \dfrac{\sum (x_i - \hat{u})^2}{n-1}$

$$\hat{\sigma}^2 = \frac{\sum (x_i - \hat{u})^2}{n}$$

Let us calculate $\lim\limits_{n \to \infty} \hat{\sigma}^2 - s^2$

$$\therefore \lim \hat{\sigma}^2 - s^2 = \lim_{n \to \infty} \left\{ \frac{\sum (x_i - \hat{u})^2}{\ldots} - \frac{\sum (x_i - \hat{u})^2}{\ldots} \right\}$$

$n \to \infty \qquad n\to\infty \Big(\quad n \qquad n-1 \quad \Big)$

$$= \lim_{n\to\infty} \left\{ \frac{(n-1) \sum (x_i - \hat{\mu})^2 - n \sum (x_i - \hat{\mu})^2}{n(n-1)} \right\}$$

$$= \lim_{n\to\infty} - \frac{\sum (x_i - \hat{\mu})^2}{n(n-1)} \longrightarrow \underline{\underline{0}} \quad \text{as } n \to \infty$$

<u>Q4.</u> <u>Given</u> :-

$$\hat{\sigma}^2_{reg} = \frac{\sum_i y_i^2 - z^T \hat{\beta}}{n-q}$$

$$\text{where } \underset{q\times 1}{z} = \sum_{i=1}^{n} \underset{q\times 1}{x_i} \underset{1\times 1}{y_i}$$

Also, $\sum (y_i - \hat{y}_i)^2 = Y^T Y - 2Y^T X \hat{\beta} + \hat{\beta} X^T X \hat{\beta}$ (given)

<u>T.P</u>:- $\hat{\beta}^T X^T X \hat{\beta} = z^T \hat{\beta}$

lets try to compute $\hat{\beta}^T X^T X \hat{\beta}$

We know that the least sq. estimator of β, $\hat{\beta} = (X^T X)^{-1} X^T Y$

$$\therefore \hat{\beta}^T X^T X \hat{\beta} = \hat{\beta}^T \underline{X^T X} \, \underline{(X^T X)}^{-1} X^T Y$$

Now here $A A^{-1} = I$, for any matrix A.

$$\therefore \hat{\beta}^T X^T X \hat{\beta} = \hat{\beta}^T I X^T Y$$

$$= \hat{\beta}^T X^T Y$$

$$= Y^T (X^T \hat{\beta}^T)^T = Y^T X \hat{\beta} \quad \text{(since its scalar matrix.)}$$

where $Y^T X = X^T Y = Z^T$

$$\therefore \hat{\beta}^T X^T X \hat{\beta} = Z^T \hat{\beta} \; //$$

(Q3) $R = D\hat{\Sigma}D$.

$$\hat{\Sigma}D = \begin{bmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{12}^2 & \cdots & \hat{\sigma}_{1p}^2 \\ \hat{\sigma}_{21}^2 & & & \\ \vdots & & & \\ \hat{\sigma}_{p1}^2 & \hat{\sigma}_{p2}^2 & \cdots & \hat{\sigma}_{pp}^2 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^{-1} & 0 & 0 & \cdots & 0 \\ 0 & & & & 0 \\ 0 & & & & \vdots \\ \vdots & & & & 0 \\ 0 & \cdots & & & \hat{\sigma}_p^{-1} \end{bmatrix}$$

Here, $\hat{\sigma}_{11}^2 \cdot \hat{\sigma}_1^{-1} = \hat{\sigma}_1$.

So we get

$$\begin{bmatrix} \hat{\sigma}_1 & \hat{\sigma}_{12}^2 \hat{\sigma}_2^{-1} & \cdots & \hat{\sigma}_{1p}^2 \hat{\sigma}_p^{-1} \\ \hat{\sigma}_{21}^2 \hat{\sigma}_1^{-1} & & & \vdots \\ \vdots & & & \vdots \\ \hat{\sigma}_{p1}^2 \hat{\sigma}_1^{-1} & \hat{\sigma}_{p2}^2 \hat{\sigma}_2^{-1} & \cdots & \hat{\sigma}_p^2 \hat{\sigma}_p^{-1} \end{bmatrix}$$

$$D\hat{\Sigma}D = \begin{bmatrix} \hat{\sigma}_1^{-1} & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ \vdots & & & & \\ 0 & & & & \hat{\sigma}_p^{-1} \end{bmatrix} \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\sigma}_1^{-1} \hat{\sigma}_1 & \dfrac{\hat{\sigma}_{12}^2}{\sigma_1 \sigma_2} & \cdots & \dfrac{\hat{\sigma}_{1p}^2}{\sigma_1 \sigma_p} \\ \vdots & & & \\ \hat{\sigma}_{p1}^2 / \sigma_p \sigma_1 & \cdots & & \hat{\sigma}_p^{-1} \hat{\sigma}_p \end{bmatrix}$$

Eeq

$$= \begin{bmatrix} 1 & \dfrac{\frac{1}{n}\sum_i x_{i,1} x_{i,1} - \bar{x}_1 \bar{x}_1}{\hat{\sigma}_1 \hat{\sigma}_2} & \cdots & \dfrac{\frac{1}{n}\sum x_{i1} x_{ip} - \bar{x}_1 \bar{x}_p}{\hat{\sigma}_1 \hat{\sigma}_p} \\ \vdots & & & \\ \dfrac{\frac{1}{n}\sum x_{ip} x_{i1} - \bar{x}_p \bar{x}_1}{\hat{\sigma}_1 \hat{\sigma}_p} & \cdots & & 1 \end{bmatrix}$$

$$r_{ijk} = \frac{\sum (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\hat{\sigma}_j \hat{\sigma}_k}$$

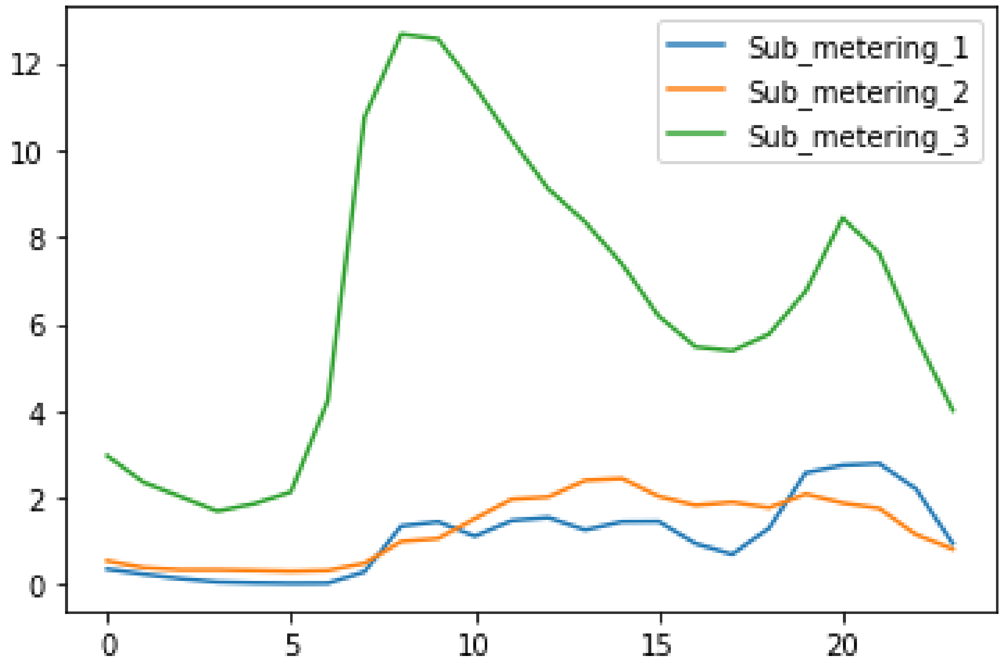
So we can write the above ~~mat~~ matrix as:

$$= \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ \vdots & & & \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} = R.$$

HW 4

Group 7 - kk1069,kas672

Computational exercise:



We can observe that in the hours 6-12, the consumption is high. We can also see that sub meter 3 has high consumption in general.

One variable Linear Regression: The closed form solution

1. Proof:
   Compute the gradient of loss function and set it to 0.

   $RSS(w_0, w_1) \ = \ \sum_1^N \left(y_i \ - \ \hat{y}_i\right)^2$

   => $\sum_1^N \left(y_i \ - \ [w_0 + w_1 x_i]\right)^2$

   $\nabla RSS(w_0) \ =- \ 2\sum_1^N \left(y_i \ - \ [w_0 + w_1 x_i]\right)$ – eq(1)

   $\nabla RSS(w_1) \ =- \ 2\sum_1^N \left(y_i \ - \ [w_0 + w_1 x_i]\right)x_i$ – eq(2)

   Set eq(1),eq(2) to 0

   We get:

$$\widehat{w}_0 = \frac{\sum_1^N y_i}{N} - \widehat{w}_1 \frac{\sum_1^N x_i}{N}$$

$$\widehat{w}_1 = \frac{\sum_1^N y_i x_i - \frac{\sum_1^N y_i \sum_1^N x_i}{N}}{\sum_1^N x_i^2 - \frac{\sum_1^N x_i \sum_1^N x_i}{N}}$$

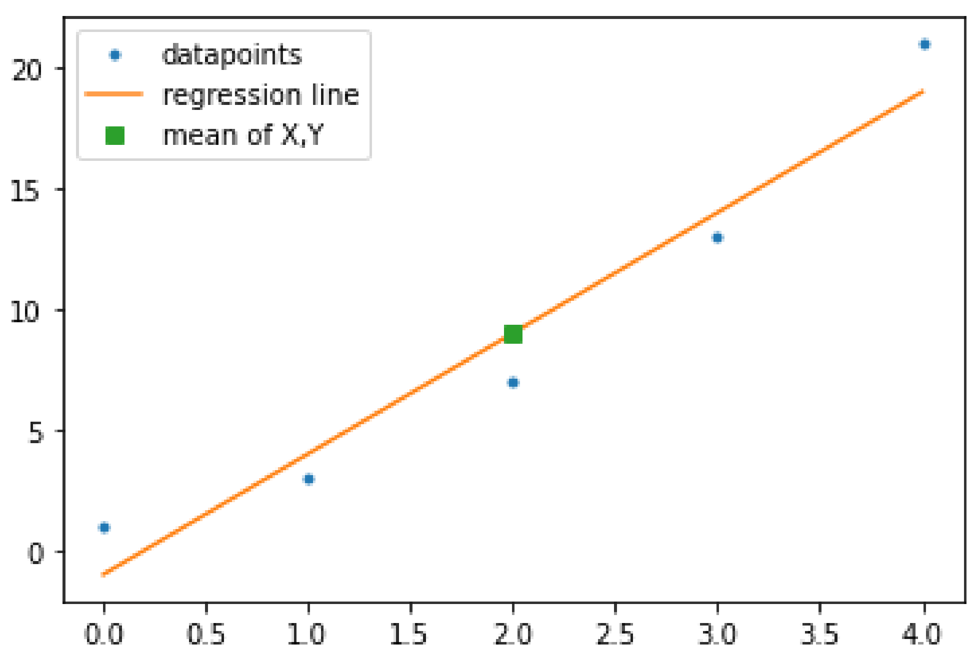
2. Interpretation:

First term: $\frac{\sum_1^N y_i}{N}$ is average house sales price

Second term: $\widehat{w}_1$ is estimate of the slope

Third term: $\frac{\sum_1^N x_i}{N}$ is average sqft of house

3. The computation is in the code file.
4. Plot:



a. The mean point i.e., $\left(\frac{\sum_1^N x_i}{N}, \frac{\sum_1^N y_i}{N}\right)$ will always lie on the regression line. This is because if we substitute this point in the line equation of the regression line $y = w_0 + w_1 x$, we get 0.

Proof:

$$\frac{\sum_1^N y_i}{N} = w_0 + w_1 \frac{\sum_1^N x_i}{N}$$

But we know $w_0 = \frac{\sum_1^N y_i}{N} - w_1 \frac{\sum_1^N x_i}{N}$ so we substitute this in the above equation, we get:

$$\frac{\sum_1^N y_i}{N} = \frac{\sum_1^N y_i}{N} - w_1 \frac{\sum_1^N x_i}{N} + w_1 \frac{\sum_1^N x_i}{N}$$

=> $\frac{\sum_1^N y_i}{N} = \frac{\sum_1^N y_i}{N}$ => LHS = RHS

So the mean point always lies on the line.

b. Even if the means of X,Y are zero, the point (0,0) will lie on the regression line.

The gradient descent:

A series of graphs are generated which show that the slope is slowly reaching 5 and the intercept is reaching -1.