Assignment-2-CH5440

CH18B020

March 13, 2022

Question-1)

Part a)

We have 4 independent variables: concentrations of CO_2, CH_4, N_2O, O_3. And, we have one independent variable: $T_{avg,\ deviation}$. We can fit a model of the form $X\beta + \beta_0 = y$ where we can estimate the parameters as,

- $\hat{\beta} = (X^T X)^{-1} X^T y$
- $\hat{\beta_0} = \bar{y} - \bar{x}^T \hat{\beta}$

The obtained model is given as:

$$T_{deviation} = 11.8 + 0.0607x_1 + 0.00591x_2 - 0.14652x_3 + 0.00804x_4 \tag{1}$$

where,

1. x_1 is concentration of CO_2
2. x_2 is concentration of CH_4
3. x_3 is concentration of N_2O
4. x_4 is concentration of O_3

The temperature deviation is positively correlated with the concentration of all gases other than N_2O, for which it is negatively correlated. This is unexpected because the correlation estimate between that and temperature deviation turns out to be positive (approx. 0.88). The coefficient values were also verified using fitlm() function in MATLAB.

Part b)

Confidence intervals:

Term	Lower Bound	Estimate	Upper Bound
Intercept	-0.2598	11.7998	23.8594
x_1	0.0336	0.0607	0.0878
x_2	0.0039	0.0059	0.0079
x_3	-0.2208	-0.1465	-0.0722
x_4	-0.0025	0.0080	0.0186

If we keep the bound for the residuals as 2, we dont see any residual greater than that value as seen in figure 1. However, the residual corresponding to the 13th data point is somewhat off positioned.

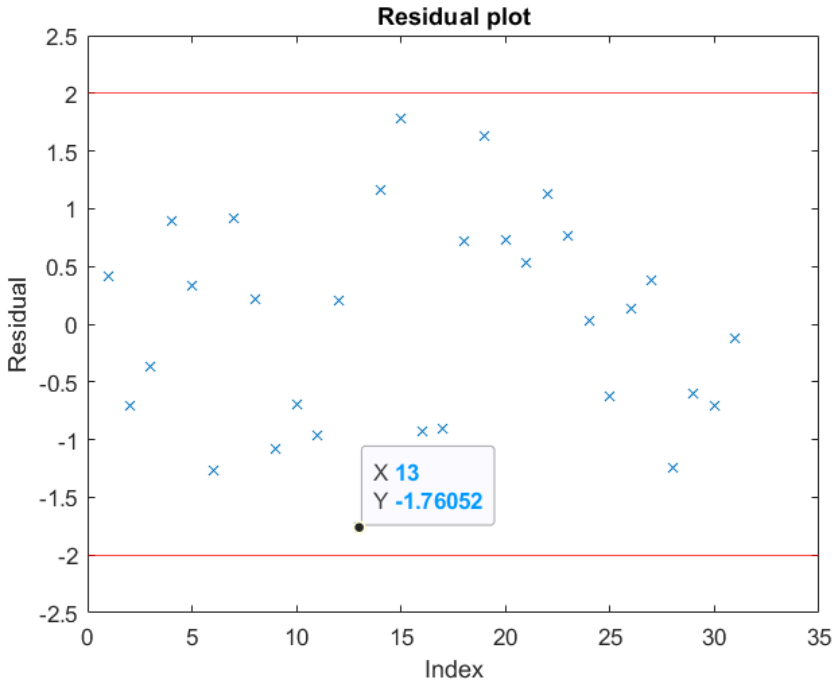

Figure 1: Residual plot for full data

Removing that point and retraining, the residuals are better now.

$$\hat{\beta_{new}} = [0.0631, 0.0066, -0.1573, 0.0086]^T \tag{2}$$

$$\hat{\beta}_{0,new} = 11.8075 \tag{3}$$

No outliers seen in the residuals of the new model.

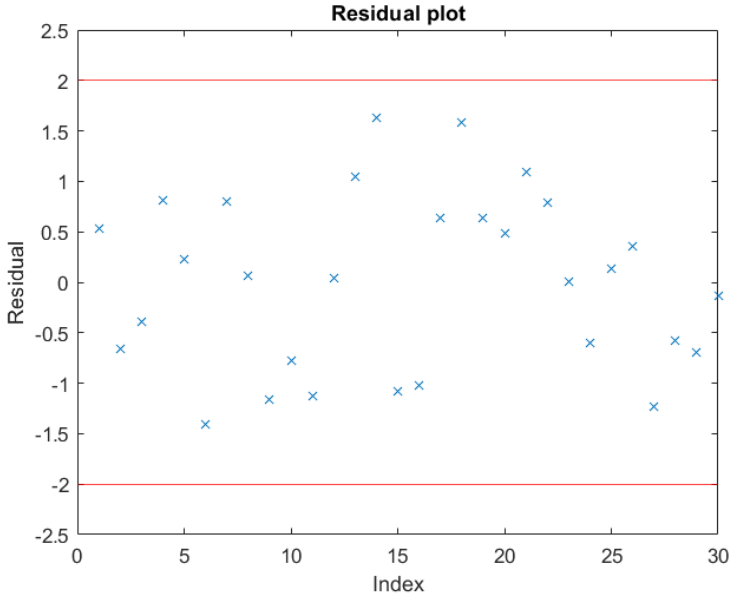

Figure 2: Residual plot after outlier removal

Part c)

fitlm() function is used to find out the pValues and standard errors.

```
mdl_c =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4

Estimated Coefficients:
                   Estimate        SE         tStat      pValue
                   ---------    ---------    -------    ---------

    (Intercept)        12.95        11.46       1.13      0.26919
    x1              0.063059     0.025766     2.4473      0.02176
    x2             0.0066088    0.0019334     3.4183    0.0021657
    x3              -0.15733      0.07074    -2.2241     0.035407
    x4             0.0086162     0.010032    0.85883      0.39859


Number of observations: 30, Error degrees of freedom: 25
Root Mean Squared Error: 0.122
R-squared: 0.858,  Adjusted R-Squared: 0.835
F-statistic vs. constant model: 37.7, p-value = 3.02e-10
```

We can see that Pvalue for concentration of Ozone coefficient is high (> 0.05) meaning it is insignificant. So we drop that first and rebuild a model.

```
mdl_c1 =


Linear regression model:
    y ~ 1 + x1 + x2 + x3

Estimated Coefficients:
                   Estimate        SE         tStat      pValue
                   ---------    ---------    -------    ---------

    (Intercept)       17.266       10.247     1.6849      0.10397
    x1              0.067786     0.025044     2.7067     0.011845
    x2             0.0064572    0.0019155     3.3709    0.0023508
    x3              -0.16813     0.069261    -2.4274     0.022435


Number of observations: 30, Error degrees of freedom: 26
Root Mean Squared Error: 0.122
R-squared: 0.854,  Adjusted R-Squared: 0.837
F-statistic vs. constant model: 50.5, p-value = 5.51e-11
```

We can see that Pvalue for intercept is high (> 0.05) meaning it is insignificant. So we drop that and rebuild the model.

```
mdl_c2 =


Linear regression model:
    y ~ x1 + x2 + x3
```

```
Estimated Coefficients:
          Estimate        SE         tStat       pValue
          ---------    ---------    -------    ----------

    x1     0.025869    0.0029784     8.6857    2.6637e-09
    x2    0.0043522    0.0015007     2.9001     0.0073299
    x3    -0.052118    0.0077743    -6.7039    3.3882e-07


Number of observations: 30, Error degrees of freedom: 27
Root Mean Squared Error: 0.126
```

We can see that Pvalue for all coefficients is very low ($<< 0.05$) which means all coefficients are highly significant.

Part d)

GWP of the gases is simply ratio of its regression coefficient (adjusted for units) with the regression coefficient of CO_2.

1. GWP of $CO_2 = 1$ (by definition)
2. GWP of $CH_4 = \frac{0.0059}{0.0607} * 10^3 = 97.405$
3. GWP of $N_2O = \frac{-0.1465}{0.0607} * 10^3 = -2413.3$

We observe that CH_4 GWP is close to the values observed over a 20 year horizon, but N_2O is not.

Question-2)

Part a)

Considering,

$$y = ln(P^{sat}) \tag{4}$$

$$x = \frac{1}{T} \tag{5}$$

we perform OLS and obtain,

$$A' = 4.7607 \tag{6}$$

$$B' = -37.896 \tag{7}$$

Part b)

The optimization problem is set similar to OLS since measurements of y is noise-free, but of course, it is nonlinear in this case.

$$\min_{\hat{P}_1^{sat},..,\hat{P}_{100}^{sat},A,B,C} \quad \sum_{i=1}^{100} (P_i^{sat} - \hat{P}_i^{sat})^2 \tag{8}$$

$$\text{s.t.} \quad ln(\hat{P}_i^{sat}) = A - \frac{B}{T_i + C} \tag{9}$$

$$\tag{10}$$

We can eliminate the equality constraint by substituting P_i^{sat} back in the objective. After doing so, using lsqnonlin() to solve the problem, one obtains,

$$A = 14.1018 \tag{11}$$

$$B = 2821.4489 \tag{12}$$

$$C = 228.7554 \tag{13}$$

Part c)

We are given that,

$$\sigma_{\epsilon_x} = 0.18 \tag{14}$$
$$\sigma_{\epsilon_y} = 2 \tag{15}$$

So we can use these values to set up a WTLS style optimization problem as given below.

$$\min_{\hat{P}_1^{sat},..,\hat{P}_{100}^{sat},\hat{T_1},...,\hat{T_{100}},A,B,C} \quad \sum_{i=1}^{100} \frac{(P_i^{sat} - \hat{P}_i^{sat})^2}{\sigma_{\epsilon_y}^2} + \frac{(T_i - \hat{T}_i)^2}{\sigma_{\epsilon_x}^2} \tag{16}$$

$$\text{s.t.} \quad ln(P_i^{sat}) = A - \frac{B}{T_i + C} \tag{17}$$

$$\tag{18}$$

To improve convergence, we have the initial guess for A,B,C as the solution of part b) (OLS problem), and the initial guess for temperature as the temperature measurement. We obtain the following estimates,

$$A = 14.1217 \tag{19}$$
$$B = 2835.2165 \tag{20}$$
$$C = 229.4130 \tag{21}$$

Part d)

Listed below are the maximum absolute error in each case.

1. Part a): 39.8234
2. Part b): 4.4844
3. Part c): 4.2112

By: S. VISHAL
CH18B020

CH5440 MVDA ASSIGNMENT-2

(3) a) $S v_1 = \lambda_1 v_1$

$$\begin{bmatrix} 7 & 21 & 34 \\ 21 & 64 & 102 \\ 34 & 102 & 186 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \\ v_{13} \end{bmatrix} = \begin{bmatrix} \lambda_1 v_{11} \\ \lambda_1 v_{12} \\ \lambda_1 v_{13} \end{bmatrix}$$

$$\Rightarrow (7-\lambda_1) v_{11} + 21 v_{12} + 34 v_{13} = 0 \quad \text{---(1)}$$

$$\Rightarrow 21 v_{11} + (64-\lambda_1) v_{12} + 102 v_{13} = 0 \quad \text{---(2)}$$

$$34 v_{11} + 102 v_{12} + (186-\lambda_1) v_{13} = 0 \quad \text{---(3)}$$

(3) $3 \times (1) - (2)$

$$-3\lambda_1 v_{11} + (\lambda_1^2 - 68\lambda_1 + 1) v_{12} = 0$$

$$\Rightarrow v_{12} = \left(\frac{3\lambda_1}{\lambda_1^2 - 68\lambda_1 + 1}\right) v_{11} \quad \text{---(4)}$$

$$v_{12} = \frac{3 \times 200.4}{199.4} v_{11} = 3.012 v_{11}$$

$34 \times (2) - 21 \times (3)$

$$\left(34(64-\lambda_1) - (21)(102)\right) v_{12} + \left(102 \times 34 - 21(186-\lambda_1)\right) v_{13} = 0$$

$$\Rightarrow v_{13} = \frac{34\lambda_1 - 34}{21\lambda_1 - 438} v_{13} \quad \text{---(5)}$$

Product of roots $= |S|$

$= 7 \times (64 \times 186 - (102)^2)$

$-21(21 \times 186 - 34 \times 102)$

$+34(102 \times 21 - 64 \times 34)$

$= 146$

$\Rightarrow \lambda_2 \lambda_3 = \frac{146}{250.4} = 0.5831$

Solving the quadratic

$$\lambda_2, \lambda_3 = \frac{+6.6 \pm \sqrt{(6.6)^2 - 4 \times 0.5831}}{2}$$

$= 6.510, 0.0896$

$\therefore \lambda_2 = 6.510, \lambda_3 = 0.0896$

Using ④ & ⑤,

$V_{22} = \left(\frac{34\lambda_2}{\lambda_2 \lambda_1 - 438}\right) V_{21}$ $\quad V_{23} = \left(\frac{34\lambda_1 - 434}{21\lambda_1 - 438}\right) V_{22}$

3.544 $\quad -2.204$

$\Rightarrow V_{22} = 2.16\, V_{21}$ $\quad V_{23} = -1.369\, V_{21}$

Normalising,

$$V_2 = \begin{bmatrix} 0.2330 \\ 0.8258 \\ -0.5135 \end{bmatrix}$$

$$\Rightarrow -v_{13} = \frac{-4820.4 + 8479.6}{-4820.4} v_{12}$$

$$\Rightarrow v_{13} = \frac{8479.6}{4820.4} \times \frac{250.4 v_{11}}{289.4}$$

$$= 5.25 v_{11}$$

Normalising,

$$v_{11} = \frac{1}{\sqrt{1 + (3.012)^2 + (5.256)^2}} = 0.1629$$

$$v_{12} = 0.4906$$

$$v_{13} = 0.8560$$

$$\therefore v_1 = \begin{bmatrix} 0.1629 \\ 0.4906 \\ 0.8560 \end{bmatrix}$$

Sum of eigen value = trace (S)

$$= 7 + 64 + 186$$

$$\Rightarrow \lambda_1 + \lambda_2 + \lambda_3 = 7 + 64 + 186$$

$$\Rightarrow \lambda_2 + \lambda_3 = 6.6$$

Once again using ④ & ⑤,

$$V_{32} = \left(\frac{3 \times 0.0896}{0.0896 - 1} \right) V_{31} = -0.2952\, V_{31}$$

$$V_{33} = \left(\frac{34 \times 0.0896 - 34}{21 \times 0.0896 - 438} \right) V_{32}$$

$$= -0.0209\, V_{32}$$

Normalising,

$$\therefore V_3 = \begin{bmatrix} 0.9589 \\ -0.2831 \\ -0.0200 \end{bmatrix}$$

Summary:

$\lambda_1 = 250.4$, $\lambda_2 = 6.5101$, $\lambda_3 = 0.0896$

$$V_1 = \begin{bmatrix} 0.1629 \\ 0.4906 \\ 0.8550 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 0.2330 \\ 0.8248 \\ -0.5135 \end{bmatrix}, \quad V_3 = \begin{bmatrix} 0.9589 \\ -0.2831 \\ -0.0200 \end{bmatrix}$$

5) If we take just first component

% variance retained = $\frac{250.4}{250.4 + 6.5 + 0.089} \times 100\%$

$= 97.436\% > 95\%$

$\therefore$ It is enough if we retain first one principal component

c) If there are 2 linear relationships, then we just pick the eigen vectors corresponding to the 2 smallest eigenvalues

such that $v^T \underline{z_s} = 0$ $\underline{z_s}$ is the shifted sample

v is the eigen vector.

So you get:

$$v_1(z_1 - \bar{z}_1) + v_2(z_2 - \bar{z}_2) + v_3(z_3 - \bar{z}_3) = 0$$

$\bar{z}_1 = 9$, $\bar{z}_2 = 68$, $\bar{z}_3 = 129$.

$$\Rightarrow v_1 z_1 + v_2 z_2 + v_3 z_3 - (v_1 \bar{z}_1 + v_2 \bar{z}_2 + v_3 \bar{z}_3) = 0$$

Substituting each of the eigen values,

Or. Rb. $0.233 z_1 + 0.8258 z_2 - 0.5135 z_3 + 7.99 = 0$

$\Rightarrow 0.233 z_1 + 0.8258 z_2 - 0.5135 z_3 + 7.9901 = 0$ ——— (6)

$0.9589 z_1 - 0.2831 z_2 - 0.02 z_3 + 13.201 = 0$ ——— (7)

where z_1 is mass

z_2 is SVL

z_3 is HLS

d) We need to project $\begin{bmatrix} 10.1 \\ 73 \\ 135.5 \end{bmatrix}$ $\begin{matrix} -9 \\ -68 \\ -129 \end{matrix}$ along the 3 different axes to get corresponding scores.

(mean shifted sample)

$$z_s = z - \frac{\bar{z}}{n} = \begin{bmatrix} 1.1 \\ 5 \\ 6.5 \end{bmatrix}$$

$v_1^T z_s$

Score along axis-1 $= v_1^T z_s = \boxed{8.18962}$

Score along axis-2 = $v_2^T z_s$ = **1.04755**

Score along axis-3 = $v_3^T z_s$ = **−0.49071**

Since there are 2 linear relationships, we can assume the last 2 scores are just due to noise. The only score that matters for compression is along axis-1 with value which has a value of 8.1962

We have 2 linear relationships in (6) & (7) and we need to estimate 2 variables. So we can just solve the eqns to get mass and HLS

Solving (6) & (7) for mass by eliminating HLS,

(6) × (+0.5135) + (7) × (−0.0201)

$\Rightarrow (0.9589 \times 0.05135) z_1 + (-0.283 \times 0.5135) z_2$

$+ (-0.020)(0.233) z_1 + (0.8259)(-0.020) z_2$

$\Rightarrow z_1 = \dfrac{-(0.8259 \times (-0.0201)) - (0.283)(0.5135)}{0.9589 \times 0.05135 - 0.0201(0.233)} \times z_3$

~~i. mass = 19.2~~

$0.4877 z_1 + 0.1618 z_2 + 6.6422 = 0$

6.6184

$\Rightarrow \text{mass} = \dfrac{0.1618 \times 73 + 6.6422}{0.4877}$

$= \boxed{10.669}$

f) Here we have 2 equations and only one variable.

So we can try to minimise the error in both equations

$\min_{\hat{z}} (z - \hat{z})^T (z - \hat{z})$

Eliminating mass from (6) & (7),

$0.8578 z_2 - 0.4877 z_3 + 4.588 = 0$

~~2.235~~

~~0.7918~~

We can set up a TLS problem so that

z_2^+, z_3^+ satisfies both (6) & (7)

& obtaining $\hat{z}_1^+$

$$\min_{z_2^+, z_3^+} (z_2 - z_2^+)^2 + (z_3 - z_3^+)^2$$

s.t. $\begin{bmatrix} 0.8578 & -0.4877 \end{bmatrix} \begin{bmatrix} z_2^+ \\ z_3^+ \end{bmatrix}$

$$= -4.586$$

$$\Rightarrow z_2^+ = \frac{-4.586 + 0.4877 z_3^+}{0.8578}$$

Subst. back we get an unconstrained quadratic problem.

$$\min_{z_2^+} (73 - z_2^+)^2 + \left(135.5 - \left(\frac{-4.586 + 0.4877 z_3^+}{0.8578}\right)\right)^2$$

$$\min_{z_3^+} \left(73 - \left(\frac{-4.586 + 0.4877 z_3^+}{0.8578}\right)\right)^2 + (135.5 - z_3^+)^2$$

$$\frac{\partial F}{\partial z_3} = 0 \Rightarrow 214.116 = 1.5685 z_3^+$$

$$\Rightarrow z_3^+ = \frac{136.062}{}$$

$$\therefore z_2^+ = 72.0115$$

$\therefore z_1^+ = 1$ (from eqn ① or ②)

$= \boxed{10\text{-}34\$\ g}$