STAT 5353: R Assignments 02

Upload your answers, as a single file, in Blackboard by 11:59 PM on Wednesday, Dec 15, 2021

PART-A

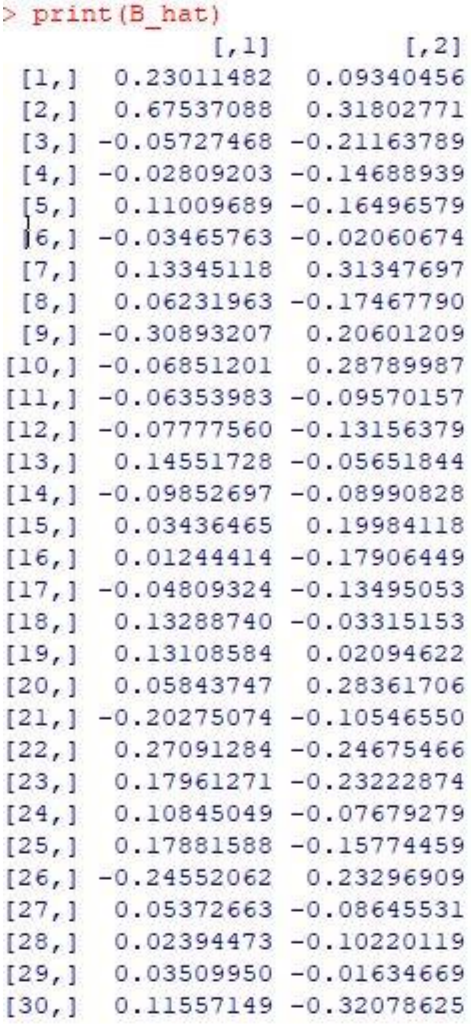
(a) Run the PLS1 regression by using 2 PLS1 predictors. You need to report the following output:
(i) $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{R}}$ matrices,
(ii) estimated regression equation using the PLS1 predictors,
(iii) value of adjusted $\boldsymbol{R}^2$ obtained after constructing 2 PLS1 predictors
Solution:

In the given Economic dataset on GDP, we consider the response variable as GDP which is the first column in our dataset and remaining columns as covariates, so there are 30 covariates which means p=30 and there are 25 different countries that are represented along rows in our dataset so, n=25.In performing this PLS1 regression, we are excluding the 5th country's data for parts (a) and (c) so, we remove the fifth row from our dataset and need to check the goodness of the fitted model by trying to predict GDP for 5th Country in parts (b) and (d). The value of n = 24, $X = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_{30} \end{bmatrix}$ and E(X) =$\mu_{P\times 1}$ and D(X) = $\Sigma_{P\times P}$ where p=30.

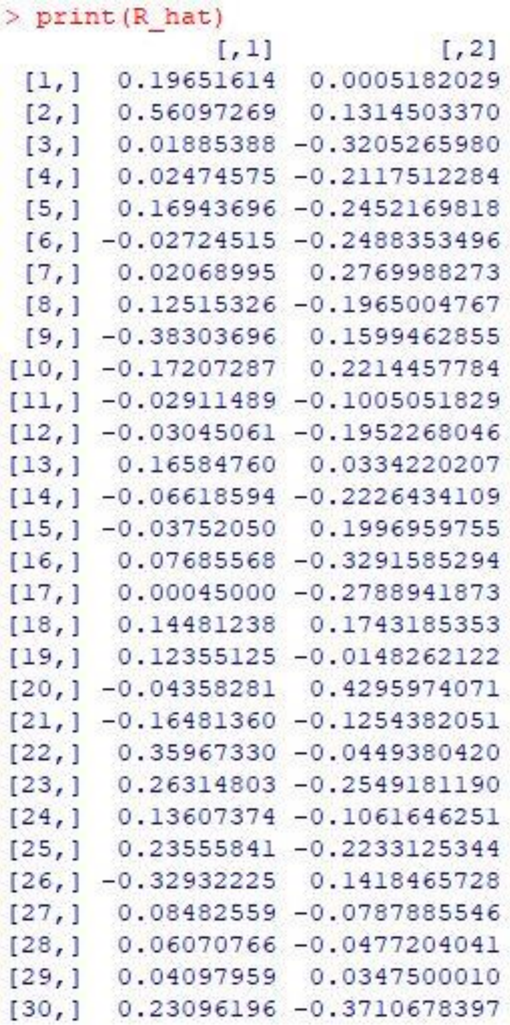
(i) Given that, we need to use 2 PLS1 predictors so, k =2. At the end 2 steps of this PLS1 regression we will have the output of $\hat{B}$ and $\hat{R}$ matrices. In general, the form of these matrices would be like:

$\hat{B}_{p\times k} = [\hat{b}_1 \hat{b}_2 \dots \hat{b}_k]$, $\hat{R}_{p\times k} \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1k} \\ \hat{r}_{21} & \hat{r}_{22} & \cdots & \hat{r}_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{r}_{p1} & \hat{r}_{p2} & \cdots & \hat{r}_{pk} \end{bmatrix}$ so here we will have,

$\hat{B}_{30\times 2} = [\hat{b}_1 \ \hat{b}_2]$, $\hat{R}_{30\times 2} \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} \\ \hat{r}_{21} & \hat{r}_{22} \\ \cdots & \cdots \\ \hat{r}_{p1} & \hat{r}_{p2} \end{bmatrix}$ and by using R, we found these matrices as shown below:



```
> print(B_hat)
              [,1]        [,2]
 [1,]  0.23011482  0.09340456
 [2,]  0.67537088  0.31802771
 [3,] -0.05727468 -0.21163789
 [4,] -0.02809203 -0.14688939
 [5,]  0.11009689 -0.16496579
 [6,] -0.03465763 -0.02060674
 [7,]  0.13345118  0.31347697
 [8,]  0.06231963 -0.17467790
 [9,] -0.30893207  0.20601209
[10,] -0.06851201  0.28789987
[11,] -0.06353983 -0.09570157
[12,] -0.07777560 -0.13156379
[13,]  0.14551728 -0.05651844
[14,] -0.09852697 -0.08990828
[15,]  0.03436465  0.19984118
[16,]  0.01244414 -0.17906449
[17,] -0.04809324 -0.13495053
[18,]  0.13288740 -0.03315153
[19,]  0.13108584  0.02094622
[20,]  0.05843747  0.28361706
[21,] -0.20275074 -0.10546550
[22,]  0.27091284 -0.24675466
[23,]  0.17961271 -0.23222874
[24,]  0.10845049 -0.07679279
[25,]  0.17881588 -0.15774459
[26,] -0.24552062  0.23296909
[27,]  0.05372663 -0.08645531
[28,]  0.02394473 -0.10220119
[29,]  0.03509950 -0.01634669
[30,]  0.11557149 -0.32078625
```



```
> print(R_hat)
              [,1]          [,2]
 [1,]  0.19651614  0.0005182029
 [2,]  0.56097269  0.1314503370
 [3,]  0.01885388 -0.3205265980
 [4,]  0.02474575 -0.2117512284
 [5,]  0.16943696 -0.2452169818
 [6,] -0.02724515 -0.2488353496
 [7,]  0.02068995  0.2769988273
 [8,]  0.12515326 -0.1965004767
 [9,] -0.38303696  0.1599462855
[10,] -0.17207287  0.2214457784
[11,] -0.02911489 -0.1005051829
[12,] -0.03045061 -0.1952268046
[13,]  0.16584760  0.0334220207
[14,] -0.06618594 -0.2226434109
[15,] -0.03752050  0.1996959755
[16,]  0.07685568 -0.3291585294
[17,]  0.00045000 -0.2788941873
[18,]  0.14481238  0.1743185353
[19,]  0.12355125 -0.0148262122
[20,] -0.04358281  0.4295974071
[21,] -0.16481360 -0.1254382051
[22,]  0.35967330 -0.0449380420
[23,]  0.26314803 -0.2549181190
[24,]  0.13607374 -0.1061646251
[25,]  0.23555841 -0.2233125344
[26,] -0.32932225  0.1418465728
[27,]  0.08482559 -0.0787885546
[28,]  0.06070766 -0.0477204041
[29,]  0.04097959  0.0347500010
[30,]  0.23096196 -0.3710678397
```

(ii) estimated regression equation using the PLS1 predictors

Solution:

The equation of fitted regression line in PLS1 regression is of the form:

$$\widehat{\boldsymbol{y}} = \overline{\boldsymbol{y}} + \widehat{\boldsymbol{q}}_{\boldsymbol{1}}\boldsymbol{t}_{\boldsymbol{1}} + \widehat{\boldsymbol{q}}_{\boldsymbol{2}}\boldsymbol{t}_{\boldsymbol{2}} + \cdots + \widehat{\boldsymbol{q}}_{\boldsymbol{k}}\boldsymbol{t}_{\boldsymbol{k}} \ = \overline{\boldsymbol{y}} + [\hat{q}_1\ \hat{q}_2 \cdots \hat{q}_k]\begin{bmatrix} t_1 \\ t_2 \\ \cdots \\ t_k \end{bmatrix} = \overline{\boldsymbol{y}} + \widehat{\boldsymbol{q}}^{\boldsymbol{T}}\boldsymbol{t}$$

By using R, we computed the values of q1_hat and t for 2 PLS1 predictors

We got,

$\overline{\boldsymbol{y}} = \boldsymbol{3}28.5714\,, \hat{q}_{2\times1} = [\hat{q}_1\ \hat{q}_2] = [22.49249\ \ 13.76378]\ and\ t1 = -0.3938637\,, t2 = -2.577551$

The estimated regression equation of this model using 2 PLS1 predictors is:

$$\widehat{\boldsymbol{y}} = \overline{\boldsymbol{y}} + \widehat{\boldsymbol{q}}_{\boldsymbol{1}}\boldsymbol{t}_{\boldsymbol{1}} + \widehat{\boldsymbol{q}}_{\boldsymbol{2}}\boldsymbol{t}_{\boldsymbol{2}} = \boldsymbol{3}28.5714 + (22.49249 * t1) + (13.76378 * t2)$$

(iii) value of adjusted $\boldsymbol{R}^{\boldsymbol{2}}$ obtained after constructing 2 PLS1 predictors

Solution: The value of adjusted R^2 obtained after constructing 2 PLS1 predictors is 0.9279999.

(b) Predict the GDP of the 5th Country (that you excluded from above model fitting) using the parameters estimated in Part (a). Compute the absolute error by using its true value
Solution:

We have to predict the GDP of the 5th country that we removed before the PLS1 regression by creating the new X as including the covariates of 5th country by creating a vector of 5th row in the dataset except GDP value in that row. The original response of 5th country is its corresponding GDP value of that row.

So, the X_new will be:

ECON_1 ECON_2 ECON_3 ECON_4 ECON_5 ECON_6 ECON_7 ECON_8 ECON_9 ECON_10
[0.231 0.799 0.188 0.845 0.771 0.301 0.391 0.348 0.885 0.398
ECON_11 ECON_12 ECON_13 ECON_14 ECON_15 ECON_16 ECON_17 ECON_18 ECON_19 ECON_20
0.119 0.578 0.985 0.875 0.377 0.860 0.483 0.185 0.794 0.703
ECON_21 ECON_22 ECON_23 ECON_24 ECON_25 ECON_26 ECON_27 ECON_28 ECON_29 ECON_30
0.630 0.181 0.644 0.843 0.417 0.186 0.563 0.867 0.500 0.103].

The original response of GDP of the 5th country from the given dataset is 345.778.

After completing 2 steps of PLS1 regression, we predict $\hat{y}^{(new)}$ as:

$$\hat{y}^{(new)} = \hat{q}t_1^{(new)} + \hat{q}_2 t_2^{(new)} = 333.1594$$

The predicted value of GDP for this 5th country using the parameters estimated in part (a) is 333.1594

The absolute error = $|\hat{y}^{(new)} - y|$ = |333.1594 – 345.778|=12.6186

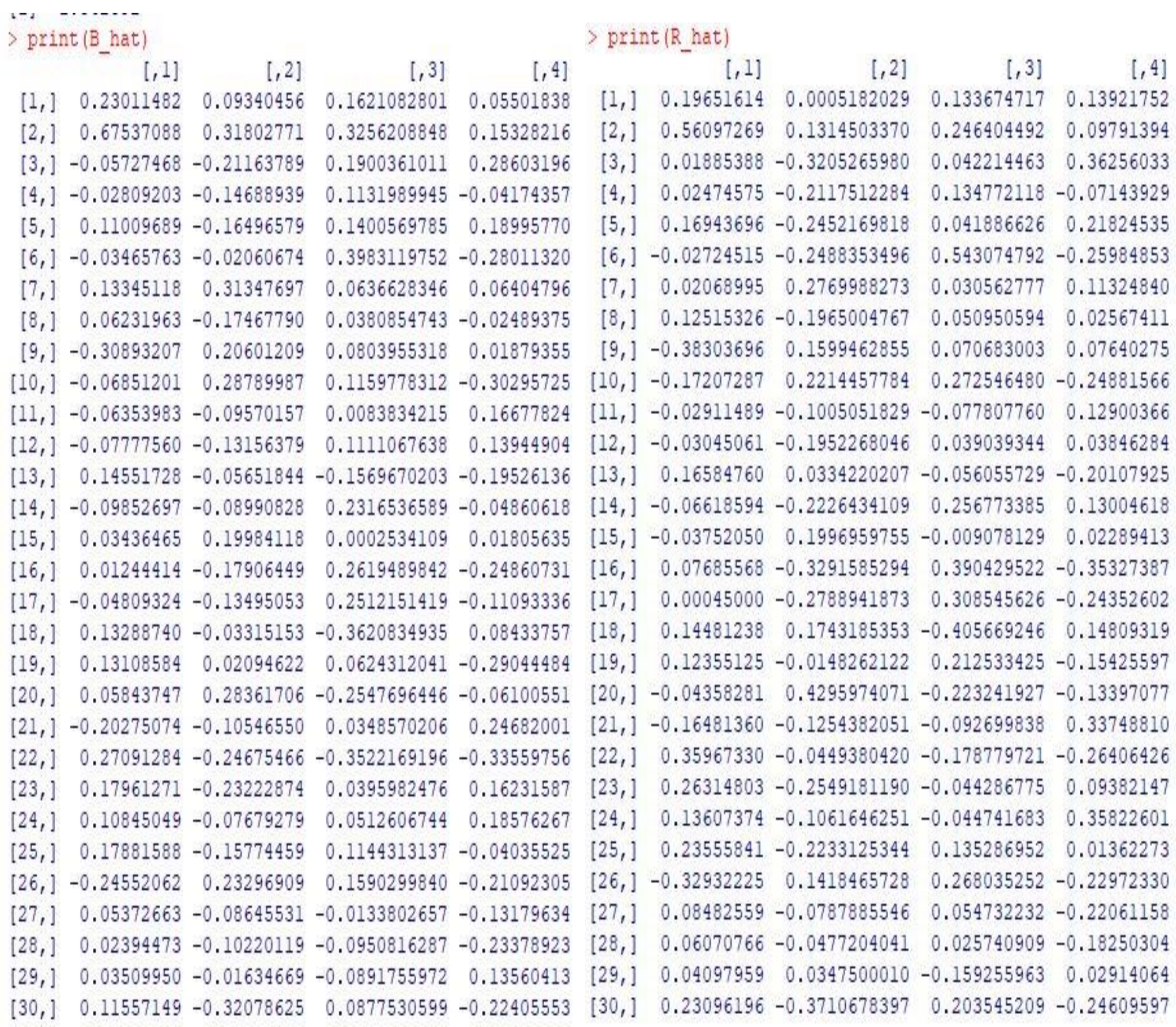
(c) Redo the analysis of Part (a) by using 4 PLS1 predictors instead of 2. Report the following output:
(i) $\hat{B}$ and $\hat{R}$ matrices,
(ii) estimated regression equation using the PLS1 predictors,
(iii) value of adjusted R^2 obtained after constructing 4 PLS1 predictors
Solution:

i) Given that, we need to use 4 PLS1 predictors so, k = 4. At the end 4 steps of this PLS1 regression we will have the output of $\hat{B}$ and $\hat{R}$ matrices. In general, the form of these matrices would be like:

$\hat{B}_{p\times k} = [\hat{b}_1\hat{b}_2 \dots \hat{b}_k]$, $\hat{R}_{p\times k}\begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1k} \\ \hat{r}_{21} & \hat{r}_{22} & \cdots & \hat{r}_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{r}_{p1} & \hat{r}_{p2} & \cdots & \hat{r}_{pk} \end{bmatrix}$ so here we will have,

$\hat{B}_{30\times 4} = [\hat{b}_1\ \hat{b}_2\ \hat{b}_3\ \hat{b}_4]$, $\hat{R}_{30\times 4}\begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \hat{r}_{13} & \hat{r}_{14} \\ \hat{r}_{21} & \hat{r}_{22} & \hat{r}_{23} & \hat{r}_{24} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{r}_{p1} & \hat{r}_{p2} & \hat{r}_{p3} & \hat{r}_{p4} \end{bmatrix}$ and by using R,

we found these matrices as shown below:



```
> print(B_hat)
             [,1]        [,2]          [,3]        [,4]
 [1,]  0.23011482  0.09340456  0.1621082801  0.05501838
 [2,]  0.67537088  0.31802771  0.3256208848  0.15328216
 [3,] -0.05727468 -0.21163789  0.1900361011  0.28603196
 [4,] -0.02809203 -0.14688939  0.1131989945 -0.04174357
 [5,]  0.11009689 -0.16496579  0.1400569785  0.18995770
 [6,] -0.03465763 -0.02060674  0.3983119752 -0.28011320
 [7,]  0.13345118  0.31347697  0.0636628346  0.06404796
 [8,]  0.06231963 -0.17467790  0.0380854743 -0.02489375
 [9,] -0.30893207  0.20601209  0.0803955318  0.01879355
[10,] -0.06851201  0.28789987  0.1159778312 -0.30295725
[11,] -0.06353983 -0.09570157  0.0083834215  0.16677824
[12,] -0.07777560 -0.13156379  0.1111067638  0.13944904
[13,]  0.14551728 -0.05651844 -0.1569670203 -0.19526136
[14,] -0.09852697 -0.08990828  0.2316536589 -0.04860618
[15,]  0.03436465  0.19984118  0.0002534109  0.01805635
[16,]  0.01244414 -0.17906449  0.2619489842 -0.24860731
[17,] -0.04809324 -0.13495053  0.2512151419 -0.11093336
[18,]  0.13288740 -0.03315153 -0.3620834935  0.08433757
[19,]  0.13108584  0.02094622  0.0624312041 -0.29044484
[20,]  0.05843747  0.28361706 -0.2547696446 -0.06100551
[21,] -0.20275074 -0.10546550  0.0348570206  0.24682001
[22,]  0.27091284 -0.24675466 -0.3522169196 -0.33559756
[23,]  0.17961271 -0.23222874  0.0395982476  0.16231587
[24,]  0.10845049 -0.07679279  0.0512606744  0.18576267
[25,]  0.17881588 -0.15774459  0.1144313137 -0.04035525
[26,] -0.24552062  0.23296909  0.1590299840 -0.21092305
[27,]  0.05372663 -0.08645531 -0.0133802657 -0.13179634
[28,]  0.02394473 -0.10220119 -0.0950816287 -0.23378923
[29,]  0.03509950 -0.01634669 -0.0891755972  0.13560413
[30,]  0.11557149 -0.32078625  0.0877530599 -0.22405553
```

```
> print(R_hat)
             [,1]          [,2]         [,3]        [,4]
 [1,]  0.19651614  0.0005182029  0.133674717  0.13921752
 [2,]  0.56097269  0.1314503370  0.246404492  0.09791394
 [3,]  0.01885388 -0.3205265980  0.042214463  0.36256033
 [4,]  0.02474575 -0.2117512284  0.134772118 -0.07143929
 [5,]  0.16943696 -0.2452169818  0.041886626  0.21824535
 [6,] -0.02724515 -0.2488353496  0.543074792 -0.25984853
 [7,]  0.02068995  0.2769988273  0.030562777  0.11324840
 [8,]  0.12515326 -0.1965004767  0.050950594  0.02567411
 [9,] -0.38303696  0.1599462855  0.070683003  0.07640275
[10,] -0.17207287  0.2214457784  0.272546480 -0.24881566
[11,] -0.02911489 -0.1005051829 -0.077807760  0.12900366
[12,] -0.03045061 -0.1952268046  0.039039344  0.03846284
[13,]  0.16584760  0.0334220207 -0.056055729 -0.20107925
[14,] -0.06618594 -0.2226434109  0.256773385  0.13004618
[15,] -0.03752050  0.1996959755 -0.009078129  0.02289413
[16,]  0.07685568 -0.3291585294  0.390429522 -0.35327387
[17,]  0.00045000 -0.2788941873  0.308545626 -0.24352602
[18,]  0.14481238  0.1743185353 -0.405669246  0.14809319
[19,]  0.12355125 -0.0148262122  0.212533425 -0.15425597
[20,] -0.04358281  0.4295974071 -0.223241927 -0.13397077
[21,] -0.16481360 -0.1254382051 -0.092699838  0.33748810
[22,]  0.35967330 -0.0449380420 -0.178779721 -0.26406426
[23,]  0.26314803 -0.2549181190 -0.044286775  0.09382147
[24,]  0.13607374 -0.1061646251 -0.044741683  0.35822601
[25,]  0.23555841 -0.2233125344  0.135286952  0.01362273
[26,] -0.32932225  0.1418465728  0.268035252 -0.22972330
[27,]  0.08482559 -0.0787885546  0.054732232 -0.22061158
[28,]  0.06070766 -0.0477204041  0.025740909 -0.18250304
[29,]  0.04097959  0.0347500010 -0.159255963  0.02914064
[30,]  0.23096196 -0.3710678397  0.203545209 -0.24609597
```

(ii) estimated regression equation using the PLS1 predictors

Solution:

The equation of fitted regression line in PLS1 regression is of the form:

$$\boldsymbol{\hat{y}} = \boldsymbol{\bar{y}} + \boldsymbol{\hat{q}_1 t_1} + \boldsymbol{\hat{q}_2 t_2} + \cdots + \boldsymbol{\hat{q}_k t_k} \;\; = \boldsymbol{\bar{y}} + [\hat{q}_1 \; \hat{q}_2 \; \cdots \hat{q}_k] \begin{bmatrix} t_1 \\ t_2 \\ \cdots \\ t_k \end{bmatrix} = \boldsymbol{\bar{y}} + \boldsymbol{\hat{q}^T t}$$

By using R, we computed the values of q1_hat and t for 4 PLS1 predictors

We got,

$\boldsymbol{\bar{y}} = \mathbf{3}28.5714$, $\hat{q}_{4\times 1} = [\hat{q}_1 \; \hat{q}_2 \; \hat{q}_3 \; \hat{q}_4] = [22.492495 \;\; 13.763776 \;\; 5.950092 \;\; 3.730687]$ and

$t1 = -0.3938637\,, t2 = -2.577551, t3 = -0.02251017$ and $t4 = -1.362892$

The estimated regression equation of this model using 4 PLS1 predictors is:

$$\boldsymbol{\hat{y}} = \boldsymbol{\bar{y}} + \boldsymbol{\hat{q}_1 t_1} + \boldsymbol{\hat{q}_2 t_2} + \boldsymbol{\hat{q}_3 t_3} + \boldsymbol{\hat{q}_4 t_4}$$

$$\boldsymbol{\hat{y}} = \mathbf{3}28.5714 + (22.49249 * t1) + (13.76378 * t2) + (5.950092 * t3) + (3.730687 * t4)$$

(iii) value of adjusted R^2 obtained after constructing 4 PLS1 predictors

Solution: The value of adjusted R^2 obtained after constructing 2 PLS1 predictors is 0.9873364.

(d) Redo all computations of Part (b) using the parameters estimated in Part (c).
Solution:
We have to predict the GDP of the 5th country that we removed before the PLS1 regression by creating the new X as including the covariates of 5th country by creating a vector of 5th row in the dataset except GDP value in that row. The original response of 5th country is its corresponding GDP value of that row.
So, the X_new will be:

ECON_1 ECON_2 ECON_3 ECON_4 ECON_5 ECON_6 ECON_7 ECON_8 ECON_9 ECON_10
[0.231 0.799 0.188 0.845 0.771 0.301 0.391 0.348 0.885 0.398
ECON_11 ECON_12 ECON_13 ECON_14 ECON_15 ECON_16 ECON_17 ECON_18 ECON_19 ECON_20
0.119 0.578 0.985 0.875 0.377 0.860 0.483 0.185 0.794 0.703
ECON_21 ECON_22 ECON_23 ECON_24 ECON_25 ECON_26 ECON_27 ECON_28 ECON_29 ECON_30
0.630 0.181 0.644 0.843 0.417 0.186 0.563 0.867 0.500 0.103].

The original response of GDP of the 5th country from the given dataset is 345.778.
After completing 4 steps of PLS1 regression, we predict $\hat{y}^{(new)}$ as:

$$\hat{y}^{(new)} = \hat{q}_1 t_1^{(new)} + \hat{q}_2 t_2^{(new)} + \hat{q}_3 t_3^{(new)} + \hat{q}_4 t_4^{(new)} = 335.1318$$

The predicted value of GDP for this 5th country using the parameters estimated in part (c) is 335.1318
The absolute error = $|\hat{y}^{(new)} - y|$ = |333.1318 – 345.778|= 10.6462

PART-B

```
## PLS1 Regression using R
## First read the data
regression_data=as.matrix(read.table("C:/MS/FALL_2021/MULTIVARIATE/R_assignments/Economic_Dataset.txt",header=TRUE))
## Read the names of variables in the data
print(colnames(regression_data))
## Set y to be the column that you want to have as your response variable
y = matrix(regression_data[-c(5) ,1],ncol=1)
## Calculate how many subjects you have, that is the value of n
n = length(y)
## Start creating the X matrix ## Include the numerical covariates
X = regression_data[-c(5),-c(1)]
X = as.matrix(X)
p = ncol(X)
sdX_hat = array(1,p)

## Now, we are starting the algorithm with 2 PLS1 predictors
##########################################################
```

```
###################################################################
k = 2
step = 0
## you need to use the following quantities later when predicting for a new observation
muX_hat = colMeans(X)
muY_hat = mean(y)
print(muY_hat)
X_temp = matrix(0,n,p)
for (j in 1:p) X_temp[ ,j] = X[ ,j] - muX_hat[j]
y_temp = y - muY_hat
## total sum of squares, useful for determining adjusted R2 later
TSS = sum(y^2) - n*((muY_hat)^2)
## create objects to store necessary outputs
B_hat = matrix(0,p,k)
R_hat = matrix(0, p, k)
q_hat = matrix(0,k,1)
Sigma_XY_Estimate_Function =
function(X_dummy,y_dummy)
{c(1/(n-1))*((t(X_dummy)-colMeans(X_dummy))%*%(y_dummy - mean(y_dummy)))}
for (step in 1:k)
{
 ## First estimate covariance between X_temp and y_temp
 Sigma_XY_hat = Sigma_XY_Estimate_Function(X_temp,y_temp)
 ## Now use that estimate to construct the b vector
 b_hat = Sigma_XY_hat/sqrt(sum(Sigma_XY_hat^2))
 ## Now use the b_vector to construct the PLS Predictor t
 t_predictor = X_temp%*%b_hat
 ## Now run p separate simple linear regressions (no intercept) to determine coefficients r_hat (vector)
 r_hat = array(0,p)

 for (j in 1:p)
 {
 out = lm(X_temp[ ,j]~0+t_predictor)
 r_hat[j] = out$coefficients
 ## update the j-th column of X_temp using residual vector from above equation
 X_temp[ ,j] = out$residuals
 }
 ## Now run one regression (no intercept) to determine coefficient q_hat (scalar)
 out = lm(y_temp~0+t_predictor)
 q_hat_value = out$coefficients
## update y_temp using residual vector from above equation
```

```
 y_temp = out$residuals
 ## Now store the necessary outputs for use in future prediction
 B_hat[ , step] = b_hat
 R_hat[ , step] = r_hat
 q_hat[step] = q_hat_value
## calculate adjusted R2
 SSE = sum(y_temp^2)
 R2 = 1 - SSE/TSS
 R2_adjusted = ((n-1)*R2 - step)/(n-step-1)
 print(R2_adjusted)
 ## t_predictors = sum(X_temp *B_hat[ ,step])
 ## print(t_predictors)
}
print(B_hat)
print(R_hat)
print(R2_adjusted)
print(q_hat)

## prediction for a future observation:

X_new = regression_data[5,-c(1) ]
print(X_new)
step = 0
## subtract training data sample mean for all variables
X_temp_new = array(0,p)
for (j in 1:p) X_temp_new[j] = X_new[j] - muX_hat[j]
y_pred_new = muY_hat
 ## initial estimate of fitted values
print(y_pred_new)
for (step in 1:k)
{
 ## Now use the appropriate column of B_hat to construct the PLS Predictor t for this new observation
 t_predictor_new = sum(X_temp_new*B_hat[ ,step])
 for (j in 1:p)
 {
 ## Now use appropriate entry of R_hat matrix to update the j-th entry of X_temp_new
 X_temp_new[j] = X_temp_new[j] - R_hat[j,step]*t_predictor_new
 }
 ## update the predicted response y_pred_new
 y_pred_new = y_pred_new + q_hat[step]*t_predictor_new
 print(y_pred_new)
```

```
}
print(y_pred_new)


################ pls algorithm for 4 predictors##############

#######################################################
#######################################################
k = 4
step = 0

muX_hat = colMeans(X)
muY_hat = mean(y)
X_temp = matrix(0,n,p)
for (j in 1:p) X_temp[ ,j] = X[ ,j] - muX_hat[j]
y_temp = y - muY_hat
TSS = sum(y^2) - n*((muY_hat)^2)

B_hat = matrix(0,p,k)
R_hat = matrix(0, p, k)
q_hat = matrix(0,k,1)
Sigma_XY_Estimate_Function =
function(X_dummy,y_dummy)
{c(1/(n-1))*((t(X_dummy)-colMeans(X_dummy))%*%(y_dummy - mean(y_dummy)))}
for (step in 1:k)
{
 Sigma_XY_hat = Sigma_XY_Estimate_Function(X_temp,y_temp)
 b_hat = Sigma_XY_hat/sqrt(sum(Sigma_XY_hat^2))
 t_predictor = X_temp%*%b_hat
 r_hat = array(0,p)

 for (j in 1:p)
 {
 out = lm(X_temp[ ,j]~0+t_predictor)
 r_hat[j] = out$coefficients
 X_temp[ ,j] = out$residuals
 }
 out = lm(y_temp~0+t_predictor)
 q_hat_value = out$coefficients
 y_temp = out$residuals
 B_hat[ , step] = b_hat
 R_hat[ , step] = r_hat
 q_hat[step] = q_hat_value
 SSE = sum(y_temp^2)
 R2 = 1 - SSE/TSS
 R2_adjusted = ((n-1)*R2 - step)/(n-step-1)
 print(R2_adjusted)
 ## t_predictors = sum(X_temp *B_hat[ ,step])
 ## print(t_predictors)
}
print(B_hat)
print(R_hat)
print(R2_adjusted)
print(q_hat)

## prediction for a future observation:

X_new = regression_data[5,-c(1) ]
print(X_new)
step = 0
X_temp_new = array(0,p)
for (j in 1:p) X_temp_new[j] = X_new[j] - muX_hat[j]
y_pred_new = muY_hat
print(y_pred_new)
for (step in 1:k)
{
 t_predictor_new = sum(X_temp_new*B_hat[ ,step])
 for (j in 1:p)
 {
 X_temp_new[j] = X_temp_new[j] - R_hat[j,step]*t_predictor_new
 }
 y_pred_new = y_pred_new + q_hat[step]*t_predictor_new
 print(y_pred_new)
}
print(y_pred_new)
```