Homework 1

Arvind Renganathan
5595189
renga016@umn.edu

October 1, 2019

1

1.1

g(x)=w$_1 x + w_0$

$\frac{\partial E}{\partial w_1} = \frac{-2}{N}\sum_{t=1}^{N}(r^t - (w_1 x^t + w_0)).x^t = 0$

$\sum_{t=1}^{N} w_0 x^t = \sum_{t=1}^{N}(r^t - w_1 x^t).x^t$

$w_0 = \frac{\sum_{t=1}^{N}(r^t - w_1 x^t}{\sum_{t=1}^{N} x^t}$ (1.1)

Similarly partially differentiating for w$_0$

$\frac{\partial E}{\partial w_0} = \frac{-2}{N}\sum_{t=1}^{N}(r^t - (w_1 x^t + w_0)).1 = 0$

$w_0 = \frac{1}{N}\sum_{t=1}^{N}(r^t - w_1 x^t)$ (1.2)

equation 1.1 and 1.2 for w$_0$

$\frac{1}{N}\sum_{t=1}^{N}(r^t - w_1 x^t) = \frac{\sum_{t=1}^{N}(r^t - w_1 x^t}{\sum_{t=1}^{N} x^t}$

$\frac{1}{N}\sum_{t=1}^{N} r^t \sum_{t=1}^{N} x^t - \frac{w_1}{N}(\sum_{t=1}^{N} x^t)^2 = \sum_{t=1}^{N} r^t x^t - w_1 \sum_{t=1}^{N}(x^t)^2$

$\sum_{t=1}^{N} r^t \sum_{t=1}^{N} x^t - w_1(\sum_{t=1}^{N} x^t)^2 = N\sum_{t=1}^{N} r^t x^t - N w_1 \sum_{t=1}^{N}(x^t)^2$

w$_1 = \frac{N\sum_{t=1}^{N} r^t x^t - \sum_{t=1}^{N} r^t \sum_{t=1}^{N} x^t}{N\sum_{t=1}^{N}(x^t)^2 - (\sum_{t=1}^{N} x^t)^2}$

Now for w_0
$w_0 = \frac{1}{N}(\sum_{t=1}^{N} r^t - w_1 \sum_{t=1}^{N} x^t$

$w_0 = \frac{1}{N}(\sum_{t=1}^{N} r^t - \sum_{t=1}^{N} x^t.(\frac{N\sum_{t=1}^{N} r^t x^t - \sum_{t=1}^{N} r^t \sum_{t=1}^{N} x^t}{N\sum_{t=1}^{N}(x^t)^2 - (\sum_{t=1}^{N} x^t)^2})$
$w_0 = \frac{1}{N}(\frac{N\sum_{t=1}^{N} r^t \sum_{t=1}^{N}(x^t)^2 - N\sum_{t=1}^{N} x^t \sum_{t=1}^{N} r^t x^t}{N\sum_{t=1}^{N}(x^t)^2 - (\sum_{t=1}^{N} x^t)^2}$

$w_0 = \frac{\sum_{t=1}^{N} r^t \sum_{t=1}^{N}(x^t)^2 - \sum_{t=1}^{N} x^t \sum_{t=1}^{N} r^t x^t}{N\sum_{t=1}^{N}(x^t)^2 - (\sum_{t=1}^{N} x^t)^2}$

1.2

$g_2(x) = v_2 x^2 + v_1 x + v_0$
$given \quad E \quad is$
$E = \frac{1}{N}\sum_{t=1}^{N}(r^t - (v_2(x^t)^2 + v_1 x^t + v_0))^2$

$Partial \quad Differentiating \quad w.r.t \quad to \quad v_2$

$\frac{\partial E}{\partial v_2} = \frac{-2}{N}\sum_{t=1}^{N}(r^t - (v_2(x^t)^2 + v_1 x^t + v_0)).(x^t)^2 = 0$

$Partial \quad Differentiating \quad w.r.t \quad to \quad v_1$

$\frac{\partial E}{\partial v_1} = \frac{-2}{N}\sum_{t=1}^{N}(r^t - (v_2(x^t)^2 + v_1 x^t + v_0)).(x^t) = 0$

$Partial \quad Differentiating \quad w.r.t \quad to \quad v_0$

$\frac{\partial E}{\partial v_2} = \frac{-2}{N}\sum_{t=1}^{N}(r^t - (v_2(x^t)^2 + v_1 x^t + v_0)).1 = 0$

$v_2\sum_{t=1}^{N}(x^t)^4 + v_1\sum_{t=1}^{N}(x^t)^3 + v_0\sum_{t=1}^{N}(x^t)^2 = \sum_{t=1}^{N} r^t(x^t)^2 \qquad (1.3)$

$$v_2\sum_{t=1}^{N}(x^t)^3 + v_1\sum_{t=1}^{N}(x^t)^2 + v_0\sum_{t=1}^{N}(x^t) = \sum_{t=1}^{N} r^t x^t \tag{1.4}$$

$$v_2\sum_{t=1}^{N}(x^t)^2 + v_1\sum_{t=1}^{N} x^t + N v_0 = \sum_{t=1}^{N} r^t \tag{1.5}$$

using 1.3 , 1.4 and 1.5 we get
$\begin{bmatrix} \sum_{t=1}^{N}(x^t)^4 & \sum_{t=1}^{N}(x^t)^3 & \sum_{t=1}^{N}(x^t)^2 \\ \sum_{t=1}^{N}(x^t)^3 & \sum_{t=1}^{N}(x^t)^2 & \sum_{t=1}^{N} x^t \\ \sum_{t=1}^{N}(x^t)^2 & \sum_{t=1}^{N} x^t & N v_0 \end{bmatrix} \begin{bmatrix} v_2 \\ v_1 \\ v_0 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N} r^t(x^t)^2 \\ \sum_{t=1}^{N} r^t x^t \\ \sum_{t=1}^{N} r^t \end{bmatrix}$
which is in the form Ax = b, where in this case x = [v2 v1 v0]^T.

1.3

Yes his claim is correct, as we are comparing training error the one for the 3 variable will always fit the model more accurately than for the 2 variables thereby leading to lesser training error or empirical loss in the 3 variable case.

1.4

No his claim is incorrect, it is the case that for a points in a linear line the model for the second case will overfit the points leading to more emprical loss for the testing set whereby the 1st model will be more accurate.

2

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 \\ 1 & 3 & 9 & 27 & 81 \\ 1 & 4 & 16 & 64 & 256 \\ 1 & 5 & 25 & 125 & 625 \end{bmatrix}$$

2.1

python Code for the answer

```
import numpy as np
matrix1 = np.matrix('[1,1,1,1,1;1, 2, 4, 8, 16;1,3,9,27,81;1,4,16,64,256;1,5,25,125,625]')

t1 = matrix1.trace()
t2 = (matrix1.transpose()).trace()
t3 = (np.matmul(matrix1.transpose(),matrix1)).trace()
t4 = (np.matmul(matrix1,matrix1.transpose())).trace()
t5 = np.dot(matrix1,matrix1.transpose()).trace()
print(t1)
print(t2)
print(t3)
print(t4)
```

tr(A) = 701
tr(A^T) = 701
$tr(A^T A)$ = 484533
$tr(AA^T)$ = 484533

2.2

From A geometric perspective a matrix represents a linear transformation, then the absolute value of the determinant represents the scaled "volume distortion" experienced by a region after being transformed. So the determinant will represent the scaled volume of the figure represented by the matrix which is basically just volume of the figure..

2.3

Without doing any computation we can clearly see that none of the rows are linear combination of the other rows in any form as they are in power series in each row thereby it implies their determinant which is |A| is not 0.

2.4

As we can see from the previous question as none of the rows are linear combination of the other rows in any form so the rank of the matrix is 5.

3

3.1 Summary of Results for K Fold

Error rates For Linear SVC with Boston 50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.35	0.16	0.27	0.43	0.06	0.23	0.14	0.19	0.21	0.48	0.25	0.13

Error rates For Linear SVC with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.29	0.31	0.17	0.14	0.53	0.35	0.16	0.04	0.24	0.14	0.24	0.13

Error rates For Linear SVC with Digit											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.08	0.13	0.04	0.04	0.04	0.07	0.10	0.14	0.05	0.11	0.08	0.03

Error rates For SVC with Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.62	0.88	0.54	0.47	0.88	0.86	0.84	0.84	0.30	0.66	0.69	0.19

Error rates For SVC with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.04	0.04	0.15	0.19	0.64	0.45	0.6	0.04	0.26	0.18	0.26	0.21

Error rates For SVC with Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.55	0.41	0.56	0.61	0.56	0.44	0.57	0.62	0.50	0.60	0.54	0.06

Error rates For Logistic Regression with Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.21	0.06	0.23	0.31	0.06	0.17	0.12	0.12	0.16	0.12	0.16	0.08

Error rates For Logistic Regression with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06	0.03	0.14	0.10	0.14	0.14	0.14	0.04	0.08	0.08	0.09	0.04

Error rates For Logistic Regression with Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06	0.14	0.04	0.02	0.03	0.05	0.09	0.10	0.05	0.09	0.07	0.04

3.2 Summary of Results for K Trainset

Error rates For Linear SVC with Boston 50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.30	0.29	0.27	0.39	0.15	0.47	0.21	0.31	0.32	0.17	0.29	0.09

Error rates For Linear SVC with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.37	0.21	0.61	0.12	0.25	0.24	0.14	0.30	0.19	0.22	0.27	0.13

Error rates For Linear SVC with Digit											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.04	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.01

Error rates For SVC with Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.16	0.22	0.22	0.17	0.18	0.23	0.13	0.20	0.17	0.15	0.19	0.03

Error rates For SVC with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.11	0.12	0.12	0.11	0.12	0.13	0.10	0.13	0.12	0.12	0.12	0.01

Error rates For SVC with Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.33	0.33	0.29	0.32	0.31	0.35	0.35	0.32	0.34	0.33	0.33	0.02

Error rates For Logistic Regression with Boston50											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.12	0.12	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.12	0.13	0.01

Error rates For Logistic Regression with Boston75											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.09	0.09	0.08	0.08	0.08	0.08	0.09	0.08	0.09	0.12	0.09	0.01

Error rates For Logistic Regression with Digits											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02	0.03	0.02	0.02	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.01

4.1 Summary of results for $\hat{X}1$(random projection) and $\hat{X}2$(quadratic projection) with K-Folds

Error rates For Linear SVC with $\hat{X}1$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.11	0.08	0.07	0.09	0.07	0.06	0.07	0.07	0.06	0.06	0.07	0.01

Error rates For Linear SVC with $\hat{X}2$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.02	0.03	0.00	0.01	0.01

Error rates For SVC with $\hat{X}1$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.43	0.36	0.33	0.41	0.29	0.35	0.41	0.31	0.36	0.33	0.36	0.04

Error rates For SVC with $\hat{X}2$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.46	0.38	0.34	0.32	0.38	0.36	0.32	0.32	0.34	0.33	0.36	0.04

Error rates For Logistic Regression with $\hat{X}1$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02	0.02	0.06	0.06	0.05	0.04	0.08	0.05	0.03	0.02	0.04	0.02

Error rates For Logistic Regression with $\hat{X}2$											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.002