SUBHODIP SAHA
5485317
saha0073@umn.edu

HW 1

1/i $g_1(x) = w_1 x + w_0$

$$E(w_1, w_0 | Z_{train}) = \frac{1}{N}\sum_{t=1}^{N}\left(r^t - (w_1 x^t + w_0)\right)^2$$

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

$$= -\frac{2}{N}\sum_{t=1}^{N} x^t(r^t - w_1 x^t - w_0) = 0 \quad \ldots (i)$$

$$= \sum_{t=1}^{N} x^t(r^t - w_1 x^t - w_0) = 0 \quad \ldots (i)$$

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

$$= \sum_{t=1}^{N}\left(r^t - (w_1 x^t + w_0)\right) = 0 \quad \ldots (ii)$$

$$= \sum_{t=1}^{N}\left(r^t - w_1 x^t - w_0\right) = 0 \quad \ldots (ii)$$

$$\Rightarrow \sum r_i - w_1 \sum x_i - w_0 N = 0$$

$$\Rightarrow \frac{\sum r_i - w_1 \sum x_i}{N} = w_0$$

$$\frac{\sum p_i - w_1 \sum x_i}{N} = w_0$$

$$\sum x_i p_i - w_1 \sum x_i^2 - w_0 \sum x_i = 0$$

$$\Rightarrow \frac{\sum x_i p_i - w_1 \sum x_i^2}{\sum x_i} = w_0$$

$$\frac{\sum p_i - w_1 \sum x_i}{N} = \frac{\sum x_i p_i - w_1 \sum x_i^2}{\sum x_i}$$

$$\Rightarrow \sum p_i \sum x_i - w_1 \left(\sum x_i\right)^2 = N \sum x_i p_i - w_1 N \sum x_i^2$$

$$\Rightarrow \sum p_i \sum x_i - N \sum x_i p_i = w_1 \left(\left(\sum x_i\right)^2 - N \sum x_i^2 \right)$$

$$\Rightarrow \boxed{w_1 = \frac{\sum p_i \sum x_i - N \sum x_i p_i}{\left(\sum x_i\right)^2 - N \sum x_i^2}} \rightarrow \text{optimal } w_1$$

$$\sum x_i p_i - w_0 \sum x_i = w_1 \sum x_i^2$$

$$\Rightarrow w_1 = \frac{\sum x_i p_i - w_0 \sum x_i}{\sum x_i^2}$$

$$\sum p_i - w_0 N = w_1 \sum x_i$$

$$\Rightarrow w_1 = \frac{\sum p_i - w_0 N}{\sum x_i}$$

$$\frac{\sum p_i - w_0 N}{\sum x_i} = \frac{\sum x_i p_i - w_0 \sum x_i}{\sum x_i^2}$$

$$\Rightarrow \sum x_i^2 \sum p_i - w_0 N \sum x_i^2 = \sum x_i \sum x_i p_i - w_0 \left(\sum x_i\right)^2$$

Scanned by CamScanner

$\Rightarrow \sum x_i^2 \sum p_i - \sum x_i \sum x_i p_i = w_0 \left(N \sum x_i^2 - (\sum x_i)^2 \right)$ 1'

$$\Rightarrow \boxed{w_0 = \frac{\sum x_i^2 \sum p_i - \sum x_i \sum x_i p_i}{N \sum x_i^2 - (\sum x_i)^2}} \rightarrow \text{optimal } w_0$$

The optimal values of $w_1 = \dfrac{\sum p_i \sum x_i - N \sum x_i p_i}{(\sum x_i)^2 - N \sum x_i^2}$

$$w_0 = \frac{\sum x_i^2 \sum p_i - \sum x_i \sum x_i p_i}{N \sum x_i^2 - (\sum x_i)^2}$$

2b

ii) $g_2(x) = v_2 x^{2020} + v_1 x + v_0$

$$E(v_2, v_1, v_0 | Z_{train}) = \frac{1}{N} \sum_{t=1}^{N} \left(p^t - (v_2 (x^t)^{2020} + v_1 x^t + v_0) \right)^2$$

$$E = \frac{1}{N} \sum \left(p_i - (v_2 x_i^{2020} + v_1 x_i + v_0) \right)^2$$

$$\frac{\partial E}{\partial v_2} = \frac{1}{N} \sum 2 \left(p_i - (v_2 x_i^{2020} + v_1 x_i + v_0) \right) (-x_i^{2020}) = 0$$

$$= -\frac{2}{N} \sum x_i^{2020} \left(p_0 - v_2 x_i^{2020} - v_1 x_i - v_0 \right) = 0$$

$$= \sum x_i^{2020} p_0 - v_2 \sum x_i^{4040} - v_1 \sum x_i^{2021} - v_0 \sum x_i^{2020} = 0$$

$$\Rightarrow v_2 \sum x_i^{4040} + v_1 \sum x_i^{2021} + v_0 \sum x_i^{2020} = \sum x_i^{2020} p_0 \quad \dots (i)$$

$\frac{\partial E}{\partial v_1} \Rightarrow$

$$E = \frac{1}{N}\sum\left(p_i - (v_2 x_i^{2020} + v_1 x_i + v_0)\right)^2$$

$$\frac{\partial E}{\partial v_1} = \frac{2}{N}\sum\left(p_i - (v_2 x_i^{2020} + v_1 x_i + v_0)\right)(-x_i)$$

$$= -\frac{2}{N}\sum x_i\left(p_i - v_2 x_i^{2020} - v_1 x_i - v_0\right) = 0$$

$$= \sum x_i p_i - v_2 \sum x_i^{2021} - v_1 \sum x_i^2 - v_0 \sum x_i = 0$$

$$\Rightarrow v_2 \sum x_i^{2021} + v_1 \sum x_i^2 + v_0 \sum x_i = \sum x_i p_i \quad \ldots (ii)$$

$$\frac{\partial E}{\partial v_3} = \frac{2}{N}\sum\left(p_i - (v_2 x_i^{2020} + v_1 x_i + v_0)\right)(-1)$$

$$= -\frac{2}{N}\sum\left(p_i - v_2 x_i^{2020} - v_1 x_i - v_0\right) = 0$$

$$\sum p_i - v_2 \sum x_i^{2020} - v_1 \sum x_i - v_0 N = 0$$

$$\Rightarrow v_2 \sum x_i^{2020} + v_1 \sum x_i + v_0 N = \sum p_i \quad \ldots (iii)$$

$\cancel{v_2 \sum x_i^{4040} + v_1}$

$$\underbrace{\begin{bmatrix} \sum x_i^{4040} & \sum x_i^{2021} & \sum x_i^{2020} \\ \sum x_i^{2021} & \sum x_i^2 & \sum x_i \\ \sum x_i^{2020} & \sum x_i & N \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} v_2 \\ v_1 \\ v_0 \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} \sum x_i^{2020} p_0 \\ \sum x_i p_i \\ \sum p_i \end{bmatrix}}_{b}$$

for solving $Av = b$,   1 2

i. I will use Gaussian elimination to make it upper triangular matrix

a. multiply first row by $-A_{i1}^{(0)}/A_{11}^{(0)}$ & add it to ith eqn.

b. ~~2nd~~ multiply 2nd row by $-A_{i2}^{(1)}/A_{22}^{(1)}$ & add it to ith eqn

So, it becomes,

$$\left[\begin{array}{ccc|c} 1 & A_{12}^{(2)} & A_{13}^{(2)} & b_1^{(2)} \\ 0 & 1 & A_{23}^{(2)} & b_2^{(2)} \\ 0 & 0 & 1 & b_3^{(2)} \end{array}\right]$$

ii. ~~Now~~ The sol^n of upper triangular matrix can be solved by backward substitution,

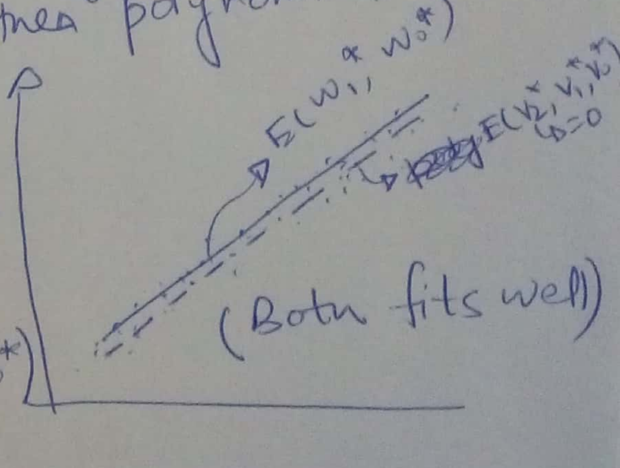
$$V_i = \frac{1}{A_{k,i}^{(2)}}\left(b_k^{(2)} - \sum_{j=i+1}^{3} A_{k,j}^{(2)} x_j \right)$$

iii. $E(v_2^*, v_1^*, v_0^* \mid Z_{train}) \leqslant E(w_1^*, w_0^* \mid Z_{train})$

Professor Gopher's claim is correct. bcz,

a. If ~~...~~ $v_2^* = 0$, ~~...~~ then the polynomial model becomes linear model.

So for example if we have linear data points both model will have same error $E(v_2^*, v_1^*, v_0^*) = E(w_1^*, w_0^*)$
$\hookrightarrow = 0$



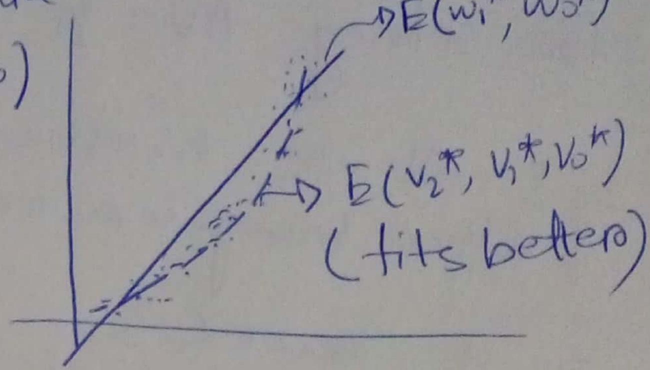
b. If we have ~~data~~ on curved data

Then $E(v_2^*, v_1^*, v_0^*) < E(w_1^*, w_0)$

~~So the Error of SE(v)~~

~~So~~, polynomial fits better.

So, Prof. Gopher's claim is correct.



2°

2 $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix}$

i. $Tr(A) = 76$

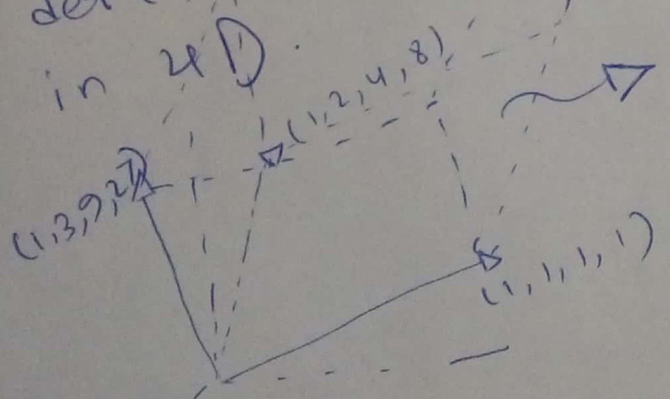
$Tr(A^T) = 76$

$Tr(A^T A) = 4182$

$Tr(A A^T) = 4182$

ii. The absolute value of $|A| = 12$ (determinant of A)

If we draw a trapezoid ~~...~~ with the row vectors of matrix A (in 4D), the det(A) will represent the volume of trapezoid in 4D.



→ The area ~~...~~ of the trapezoid will be equal to det(A).

(I have drawn in 3D)

iii) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix}$

Let's write

$$\begin{bmatrix} 1 \\ 4 \\ 16 \\ \hline 64 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \hline 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 2 \\ 4 \\ \hline 8 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 3 \\ 9 \\ \hline 27 \end{bmatrix}$$

$$1 = \alpha_1 + \alpha_2 + \alpha_3$$
$$4 = \alpha_1 + 2\alpha_2 + 3\alpha_3$$
$$16 = \alpha_1 + 4\alpha_2 + 9\alpha_3$$

Solving $\alpha_1, \alpha_2, \alpha_3$ we get
$\alpha_1 = 1, \alpha_2 = -3, \alpha_3 = 3$

Now 4th eq^n
$\alpha_1 + 8\alpha_2 + 27\alpha_3 = 1 - 24 + 81 = 58 \neq 64$

So, (we can't have solution of $\alpha_1, \alpha_2, \alpha_3$)

So, all row vectors are linearly independent.

Problem 3 (part i)

1. Error rates for linear SVC with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.4	0.28	0.08	0.12	0.18	0.34	0.16	0.1	0.28	0.22	0.21600000000000003	0.10150862032359617

2. Error rates for linear SVC with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.28	0.52	0.06	0.48	0.32	0.12	0.14	0.06	0.1	0.14	0.22200000000000003	0.16110865898517063

3. Error rates for linear SVC with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06703911	0.01675978	0.06703911	0.05586592	0.05586592	0.04469274	0.05586592	0.03910615	0.03910615	0.03910615	0.04804469273743019	0.014610834447622378

4. Error rates for SVC with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.18	0.2	0.26	0.22	0.22	0.32	0.18	0.26	0.22	0.28	0.23399999999999999	0.042941821107167766

5. Error rates for SVC with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.18	0.12	0.28	0.28	0.3	0.26	0.22	0.34	0.28	0.12	0.23800000000000004	0.07180529228406497

6. Error rates for SVC with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00558659	0.00558659	0.00558659	0.00558659	0.03351955	0.00558659	0.02234637	0.01117318	0.	0.01117318	0.010614525139664821	0.00949720670391059

7. Error rates for Logistic Regression with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.12	0.1	0.18	0.08	0.18	0.2	0.16	0.18	0.18	0.12	0.15000000000000005	0.03924283374069719

8. Error rates for Logistic Regression with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.14	0.04	0.14	0.14	0.08	0.14	0.06	0.12	0.1	0.1	0.10600000000000001	0.034698703145794936

9. Error rates for Logistic Regression with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02793296	0.05027933	0.02793296	0.02234637	0.04469274	0.02793296	0.06703911	0.02793296	0.02793296	0.02234637	0.03463687150837989	0.013865557146358502

Problem 3 (part ii)

1. Error rates for linear SVC with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.34920635	0.22222222	0.13492063	0.37301587	0.13492063	0.36507937	0.25396825	0.22222222	0.34920635	0.14285714	0.25476190476190474	0.09346564964947868

2. Error rates for linear SVC with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.15873016	0.38888889	0.11111111	0.21428571	0.62698413	0.26984127	0.15873016	0.13492063	0.16666667	0.3015873	0.25317460317460316	0.14892112217506817

3. Error rates for linear SVC with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.06013363	0.077951	0.06458797	0.05122494	0.0467706	0.05122494	0.06458797	0.05345212	0.0467706	0.077951	0.0594654788418708	0.011093460494799343

4. Error rates for SVC with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.19047619	0.23015873	0.26984127	0.21428571	0.23015873	0.22222222	0.26190476	0.28571429	0.26190476	0.23809524	0.24047619047619045	0.027504322938354694

5. Error rates for SVC with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.23015873	0.26190476	0.22222222	0.24603175	0.24603175	0.23015873	0.26984127	0.19047619	0.19047619	0.23809524	0.23253968253968252	0.025109987332629174

6. Error rates for SVC with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00890869	0.01336303	0.00890869	0.01113586	0.00222717	0.01336303	0.00890869	0.00668151	0.01113586	0.01781737	0.010244988864142523	0.004008908685968835

7. Error rates for Logistic Regression with Boston50

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.17460317	0.16666667	0.15079365	0.13492063	0.18253968	0.15079365	0.14285714	0.14285714	0.11111111	0.15079365	0.15079365079365079	0.01944039478399347

8. Error rates for Logistic Regression with Boston75

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.13492063	0.13492063	0.08730159	0.11904762	0.15079365	0.11904762	0.14285714	0.14285714	0.0952381	0.07936508	0.12063492063492065	0.0240202316673358

9. Error rates for Logistic Regression with Digits

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.04008909	0.03340757	0.04231626	0.03786192	0.03563474	0.04008909	0.04454343	0.0467706	0.04008909	0.04008909	0.04008908685968817	0.00372677071952817

Problem 4

1. Error rates for LinearSVC with X1

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.0726257	0.05586592	0.06703911	0.10055866	0.12290503	0.07821229	0.10055866	0.06145251	0.06703911	0.16759777	0.08938547486033521	0.032861321049950265

2. Error rates for Linear SVC with X2

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00558659	0.01117318	0.0	0.04469274	0.0	0.01117318	0.01117318	0.03351955	0.01117318	0.0	0.012849162011173176	0.014144121677287314

3. Error rates for SVC with X1

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02234637	0.01675978	0.03351955	0.01675978	0.01675978	0.00558659	0.02234637	0.01675978	0.01117318	0.00558659	0.016759776536312866	0.00790063442666531

4. Error rates for SVC with X2

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.00558659	0.01117318	0.0	0.01117318	0.011173	0.0	0.02793296	0.0	0.02234637	0.01117318	0.01005586592178771	0.008938547486033517

5. Error rates for LogisticRegression with X1

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.05586592	0.05027933	0.06145251	0.05027933	0.07821229	0.07821229	0.04469274	0.05027933	0.08379888	0.08379888	0.06368715083798883	0.014822904090973853

6. Error rates for LogisticRegression with X2

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean	SD
0.02793296	0.00558659	0.01117318	0.00558659	0.01117318	0.01117318	0.0	0.02234637	0.02234637	0.01117318	0.012849162011173188	0.008305066339284073