ML assignment 2

Saurabh Vaidya
Email- vaidya.saur@husky.neu.edu

February 2020

1 Problem 1

1.1 a

For Naive Bayes

$$Y = \begin{cases} +, & \text{if } p(y=+|x) \geq p(y=-|x) \\ -, & \text{otherwise} \end{cases} \tag{1}$$

For X = (-1,-1)

$$\begin{aligned} p(y=+|x) &= \prod_{j=1}^{d} p(x_j|y=+)p(+) && (2)\\ &= p(x_1=-1|y=+)p(x_2=-1|y=+)p(+) && (3)\\ &= 1/2*1/2*1/2 && (4)\\ &= 1/8 && (5)\\ p(y=-|x) &= \prod_{j=1}^{d} p(x_j|y=-)p(-) && (6)\\ &= p(x_1=-1|y=-)p(x_2=-1|y=-)p(-) && (7)\\ &= 1/2*1/2*1/2 && (8)\\ &= 1/8 && (9)\\ y &= + && (10)\\ & && (11) \end{aligned}$$

For X = (-1,1)

$$\begin{aligned} p(y=+|x) &= \prod_{j=1}^{d} p(x_j|y=+)p(+) && (12)\\ &= p(x_1=-1|y=+)p(x_2=1|y=+)p(+) && (13)\\ &= 1/2*1/2*1/2 && (14)\\ &= 1/8 && (15)\\ p(y=-|x) &= \prod_{j=1}^{d} p(x_j|y=-)p(-) && (16)\\ &= p(x_1=-1|y=-)p(x_2=1|y=-)p(-) && (17)\\ &= 1/2*1/2*1/2 && (18)\\ &= 1/8 && (19)\\ y &= + && (20)\\ & && (21) \end{aligned}$$

For X = (1,-1)

$$
\begin{aligned}
p(y=+|x) &= \prod_{j=1}^{d} p(x_j|y=+)p(+) && (22)\\
&= p(x_1=1|y=+)p(x_2=-1|y=+)p(+) && (23)\\
&= 1/2 * 1/2 * 1/2 && (24)\\
&= 1/8 && (25)\\
p(y=-|x) &= \prod_{j=1}^{d} p(x_j|y=-)p(-) && (26)\\
&= p(x_1=1|y=-)p(x_2=-1|y=-)p(-) && (27)\\
&= 1/2 * 1/2 * 1/2 && (28)\\
&= 1/8 && (29)\\
y &= + && (30)\\
& && (31)
\end{aligned}
$$

For X = (1,1)

$$
\begin{aligned}
p(y=+|x) &= \prod_{j=1}^{d} p(x_j|y=+)p(+) && (32)\\
&= p(x_1=1|y=+)p(x_2=1|y=+)p(+) && (33)\\
&= 1/2 * 1/2 * 1/2 && (34)\\
&= 1/8 && (35)\\
p(y=-|x) &= \prod_{j=1}^{d} p(x_j|y=-)p(-) && (36)\\
&= p(x_1=1|y=-)p(x_2=1|y=-)p(-) && (37)\\
&= 1/2 * 1/2 * 1/2 && (38)\\
&= 1/8 && (39)\\
y &= + && (40)\\
& && (41)
\end{aligned}
$$

Accuracy Calculation

$$
\begin{aligned}
\text{Accuracy} &= \frac{\text{No of correct predictions}}{No of total samples} && (42)\\
&= \frac{2}{4} && (43)\\
&= 50\% && (44)
\end{aligned}
$$

1.2 b

The new input space is

+1	x_1	x_2	$x_1 x_2$	x_1^2	x_2^2	y
1	-1	-1	1	1	1	-
1	-1	1	-1	1	1	+
1	1	-1	-1	1	1	+
1	1	1	1	1	1	-

For X = (1,-1,-1,1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{45}$$
$$= p(+1|y=+)p(x_1=-1|y=+)p(x_2=-1|y=+)p(x_1x_2=1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+)p(+) \tag{46}$$
$$= 1*1/2*1/2*0*1*1 \tag{47}$$
$$= 0 \tag{48}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{49}$$
$$= p(+1|y=-)p(x_1=-1|y=-)p(x_2=-1|y=-)p(x_1x_2=1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-)p(-) \tag{50}$$
$$= 1*1/2*1/2*1*1*1 \tag{51}$$
$$= 1/4 \tag{52}$$
$$y = - \tag{53}$$
$$\tag{54}$$

For X = (1,-1,1,-1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{55}$$
$$= p(+1|y=+)p(x_1=-1|y=+)p(x_2=1|y=+)p(x_1x_2=-1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+)p(+) \tag{56}$$
$$= 1*1/2*1/2*1*1*1 \tag{57}$$
$$= 1/4 \tag{58}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{59}$$
$$= p(+1|y=-)p(x_1=-1|y=-)p(x_2=1|y=-)p(x_1x_2=-1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-)p(-) \tag{60}$$
$$= 1*1/2*1/2*0*1*1 \tag{61}$$
$$= 0 \tag{62}$$
$$y = + \tag{63}$$
$$\tag{64}$$

For X = (1,1,-1,-1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{65}$$
$$= p(+1|y=+)p(x_1=1|y=+)p(x_2=-1|y=+)p(x_1x_2=-1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+)p(+) \tag{66}$$
$$= 1*1/2*1/2*1*1*1 \tag{67}$$
$$= 1/4 \tag{68}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{69}$$
$$= p(+1|y=-)p(x_1=1|y=-)p(x_2=-1|y=-)p(x_1x_2=-1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-)p(-) \tag{70}$$
$$= 1*1/2*1/2*0*1*1 \tag{71}$$
$$= 0 \tag{72}$$
$$y = + \tag{73}$$
$$\tag{74}$$

For X = (1,1,1,1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{75}$$
$$= p(+1|y=+)p(x_1=1|y=+)p(x_2=1|y=+)p(x_1x_2=1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+)p(+) \tag{76}$$
$$= 1*1/2*1/2*0*1*1 \tag{77}$$
$$= 0 \tag{78}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{79}$$
$$= p(+1|y=-)p(x_1=-1|y=-)p(x_2=-1|y=-)p(x_1x_2=1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-)p(-) \tag{80}$$
$$= 1*1/2*1/2*1*1*1 \tag{81}$$
$$= 1/4 \tag{82}$$
$$y = - \tag{83}$$
$$\tag{84}$$

Accuracy Calculation

$$\text{Accuracy} = \frac{\text{No of correct predictions}}{No of total samples} \tag{85}$$
$$= \frac{4}{4} \tag{86}$$
$$= 100\% \tag{87}$$

1.3 c

The new input space is

+1	x_1	x_2	x_1x_2	x_1^2	x_2^2	$-x_1x_2$	y
1	-1	-1	1	1	1	-1	-
1	-1	1	-1	1	1	1	+
1	1	-1	-1	1	1	1	+
1	1	1	1	1	1	-1	-

For X = (1,-1,-1,1,1,1,-1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{88}$$
$$= p(+1|y=+)p(x_1=-1|y=+)p(x_2=-1|y=+)p(x_1x_2=1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+) \tag{89}$$
$$p(-x_1x_2=-1|y=+)p(+) \tag{90}$$
$$= 1*1/2*1/2*0*1*1*0 \tag{91}$$
$$= 0 \tag{92}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{93}$$
$$= p(+1|y=-)p(x_1=-1|y=-)p(x_2=-1|y=-)p(x_1x_2=1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-) \tag{94}$$
$$p(-x_1x_2=-1|y=-)p(-) \tag{95}$$
$$= 1*1/2*1/2*1*1*1*1 \tag{96}$$
$$= 1/4 \tag{97}$$
$$y = - \tag{98}$$
$$\tag{99}$$

For X = (1,-1,1,-1,1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{100}$$
$$= p(+1|y=+)p(x_1=-1|y=+)p(x_2=1|y=+)p(x_1x_2=-1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+) \tag{101}$$
$$p(-x_1x_2=1|y=+)p(+) \tag{102}$$
$$= 1*1/2*1/2*1*1*1*1 \tag{103}$$
$$= 1/4 \tag{104}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{105}$$
$$= p(+1|y=-)p(x_1=-1|y=-)p(x_2=1|y=-)p(x_1x_2=-1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-) \tag{106}$$
$$p(-x_1x_2=1|y=-)p(-) \tag{107}$$
$$= 1*1/2*1/2*0*1*1*0 \tag{108}$$
$$= 0 \tag{109}$$
$$y = + \tag{110}$$
$$\tag{111}$$

For X = (1,1,-1,-1,1,1,1)

$$p(y=+|x) = \prod_{j=1}^{d} p(x_j|y=+)p(+) \tag{112}$$
$$= p(+1|y=+)p(x_1=1|y=+)p(x_2=-1|y=+)p(x_1x_2=-1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+) \tag{113}$$
$$p(-x_1x_2=1|y=+)p(+) \tag{114}$$
$$= 1*1/2*1/2*1*1*1*1 \tag{115}$$
$$= 1/4 \tag{116}$$
$$p(y=-|x) = \prod_{j=1}^{d} p(x_j|y=-)p(-) \tag{117}$$
$$= p(+1|y=-)p(x_1=1|y=-)p(x_2=-1|y=-)p(x_1x_2=-1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-) \tag{118}$$
$$p(-x_1x_2=1|y=-)p(-) \tag{119}$$
$$= 1*1/2*1/2*0*1*1*0 \tag{120}$$
$$= 0 \tag{121}$$
$$y = + \tag{122}$$
$$\tag{123}$$

For X = (1,1,1,1,1,1,-1)

$$
\begin{aligned}
p(y=+|x) &= \prod_{j=1}^{d} p(x_j|y=+)p(+) && (124)\\
&= p(+1|y=+)p(x_1=1|y=+)p(x_2=1|y=+)p(x_1x_2=1|y=+)p(x_1^2=1|y=+)p(x_2^2=1|y=+) && (125)\\
& p(-x_1x_2=-1|y=+)p(+) && (126)\\
&= 1*1/2*1/2*0*1*1*0 && (127)\\
&= 0 && (128)\\
p(y=-|x) &= \prod_{j=1}^{d} p(x_j|y=-)p(-) && (129)\\
&= p(+1|y=-)p(x_1=-1|y=-)p(x_2=-1|y=-)p(x_1x_2=1|y=-)p(x_1^2=1|y=-)p(x_2^2=1|y=-) && (130)\\
& p(-x_1x_2=-1|y=-)p(-) && (131)\\
&= 1*1/2*1/2*1*1*1*1 && (132)\\
&= 1/4 && (133)\\
y &= - && (134)\\
& && (135)
\end{aligned}
$$

Accuracy Calculation

$$
\begin{aligned}
\text{Accuracy} &= \frac{\text{No of correct predictions}}{No of total samples} && (136)\\
&= \frac{4}{4} && (137)\\
&= 100\% && (138)
\end{aligned}
$$

Adding this dependent feature had no impact on output from previous problem.
This feature helps the model to attain perfect accuracy.

2 Problem 2

2.1 a

For a point on the optimal decision surface,

$$P(Y=1|x) = P(Y=0|x) \tag{139}$$

$$\frac{e^{\frac{-(x-m_1)^T\Sigma_1^{-1}(x-m_1)}{2}}}{2\pi|\Sigma_1|^{1/2}}.P(Y=1) = \frac{e^{\frac{-(x-m_0)^T\Sigma_0^{-1}(x-m_0)}{2}}}{2\pi|\Sigma_0|^{1/2}}.P(Y=0) \tag{140}$$

$$\tag{141}$$

$$\text{Since P(Y=0)=P(Y=1) and } \Sigma_0 = \Sigma_1 = I_2 \tag{142}$$

$$e^{\frac{-(x-m_0)^T I_2(x-m_0)}{2}} = e^{\frac{-(x-m_1)^T I_2(x-m_1)}{2}} \tag{143}$$

$$log(e^{\frac{-(x-m_0)^T I_2(x-m_0)}{2}}) = log(e^{\frac{-(x-m_1)^T I_2(x-m_1)}{2}}) \tag{144}$$

$$(x-m_0)^T I_2(x-m_0) = (x-m_1)^T I_2(x-m_1) \tag{145}$$

$$(x^T - m_0^T)(x-m_0) = (x^T - m_1^T)(x-m_1) \tag{146}$$

$$x^Tx - x^Tm_0 - m_0^Tx + m_0Tm_0 = x^Tx - x^Tm_1 - m_1^Tx + m_1^Tm_1 \tag{147}$$

$$-x^Tm_0 + x^Tm_1 + m_1^Tx - m_0^Tx + m_0^Tm_0 - m_1^Tm_1 = 0 \tag{148}$$

$$x^T(m_1 - m_0) + (m_1^T - m_0^T)x + m_0^Tm_0 - m_1^Tm_1 = 0 \tag{149}$$

$$x^T\begin{bmatrix}5\\1\end{bmatrix} + \begin{bmatrix}5 & 1\end{bmatrix}x + 5 - 45 = 0 \tag{150}$$

$$\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}5\\1\end{bmatrix} + \begin{bmatrix}5 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} - 40 = 0 \tag{151}$$

$$5x_1 + x_2 + 5x_1 + x_2 - 40 = 0 \tag{152}$$

$$10x_1 + 2x_2 - 40 = 0 \tag{153}$$

$$5x_1 + x_2 - 20 = 0 \tag{154}$$

$$\tag{155}$$

2.2 c

$$\frac{1}{2\pi|\Sigma_1|^{1/2}}e^{\frac{-(x-m_1)^T\Sigma_1^{-1}(x-m_1)}{2}}.P(Y=1) = \frac{1}{2\pi|\Sigma_0|^{1/2}}e^{\frac{-(x-m_0)^T\Sigma_0^{-1}(x-m_0)}{2}}.P(Y=0) \tag{156}$$

$$\frac{-1}{2}(x-m_1)^T\Sigma_1^{-1}(x-m_1) + log(P(Y=1)) - \frac{1}{2}log(|\Sigma_1|) = \frac{-1}{2}(x-m_0)^T\Sigma_0^{-1}(x-m_0) + log(P(Y=0)) - \frac{1}{2}log(|\Sigma_0|) \tag{157}$$

$$(x^T\Sigma_0^{-1} - m_1^T\Sigma_0^{-1})(x-m_0) - (x^T\Sigma_1^{-1} - m_1^T\Sigma_1^{-1})(x-m_1) = 2log(P(Y=1)) - 2log(P(Y=0)) + log(|\Sigma_1|) - log(|\Sigma_0|) \tag{158}$$

$$(x^T\Sigma_0^{-1}x - x^T\Sigma_0^{-1}m_0 - m_0^T\Sigma_0^{-1}x + m_0^T\Sigma_0^{-1}m_0) - (x^T\Sigma_1^{-1}x - x^T\Sigma_1^{-1}m_1 - m_1^T\Sigma_1^{-1}x + m_1^T\Sigma_1^{-1}m_1) = 2log\left[\frac{P(Y=1)}{P(Y=0)}\right] + log\left[\frac{|\Sigma_1|}{|\Sigma_0|}\right] \tag{159}$$

$$x^T\Sigma_0^{-1}x - x^T\Sigma_1^{-1}x - x^T\Sigma_0^{-1}m_0 + x^T\Sigma_1^{-1}m_1 - m_0^T\Sigma_0^{-1}x + m_1^T\Sigma_1^{-1}x + m_0^T\Sigma_0^{-1}m_0 + m_1^T\Sigma_1^{-1}m_1 = 2log\left[\frac{P(Y=1)}{P(Y=0)}\right] + log\left[\frac{|\Sigma_1|}{|\Sigma_0|}\right] \tag{160}$$

$$x^T\Sigma_0^{-1}x - x^T\Sigma_1^{-1}x - x^T(\Sigma_0^{-1}m_0 - \Sigma_1^{-1}m_1) - (m_0^T\Sigma_0^{-1} + m_1^T\Sigma_1^{-1})x + m_0^T\Sigma_0^{-1}m_0 + m_1^T\Sigma_1^{-1}m_1 = 2log\left[\frac{P(Y=1)}{P(Y=0)}\right] + log\left[\frac{|\Sigma_1|}{|\Sigma_0|}\right] \tag{161}$$

$$\tag{162}$$

The shape of the optimal surface is a polynomial with degree 2 since the equation is quadratic.
In part a and b as will be seen below, the shape of the optimal decision surface becomes a hyperplane/plane since the equation is linear

2.3 b

From the above equation in part c when $\Sigma_0 = \Sigma_1 = \sigma^2 I_2$

$$\frac{-x^T(m_0 - m_1)}{\sigma^2} - \frac{(m_0^T - m_1^T)x}{\sigma^2} + \frac{m_0^T m_0}{\sigma^2} + \frac{m_1^T m_1}{\sigma^2} = 2log\left[\frac{P(Y=1)}{P(Y=0)}\right] \tag{163}$$

$$-x^T(m_0 - m_1) - (m_0^T - m_1^T)x + m_0^T m_0 + m_1^T m_1 = 2\sigma^2 log\left[\frac{P(Y=1)}{P(Y=0)}\right] \tag{164}$$

$$\tag{165}$$

where $m_0 = (m_{01}, m_{02})$ and $m_1 = (m_{11}, m_{12})$

3 Problem 3

3.1 a

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\sum_{j=0}^{d}(w_j x_{ij} - y_i)^2 \tag{166}$$

$$\frac{\partial e_i^2}{\partial w_0} = 2\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{i0} \tag{167}$$

$$\frac{\partial e_i^2}{\partial w_1} = 2\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{i1} \tag{168}$$

$$\vdots$$

$$\frac{\partial e_i^2}{\partial w_d} = 2\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{id} \tag{169}$$

$$\tag{170}$$

The gradient descent updates would be

$$w_0^{t+1} = w_0^t - 2\alpha\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{i0} \tag{171}$$

$$w_1^{t+1} = w_1^t - 2\alpha\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{i1} \tag{172}$$

$$\vdots$$

$$w_d^{t+1} = w_d^t - 2\alpha\sum_{i=1}^{n}(\sum_{j=0}^{d}(w_j x_{ij} - y_i))x_{id} \tag{173}$$

$$\tag{174}$$

3.2 b

$$
\begin{aligned}
\sum_{i=1}^{n} r_i^2 &= \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2}{\sum_{j=1}^{d} w_j^2} && (175)\\
\frac{\partial r_i^2}{\partial w_0} &= \frac{2(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)}{\sum_{j=1}^{d} w_j^2} && (176)\\
\frac{\partial r_i^2}{\partial w_1} &= \frac{-(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 2w_1}{(\sum_{j=1}^{d} w_j^2)^2} + \frac{2(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{i1})}{\sum_{j=1}^{d} w_j^2} && (177)\\
&= \frac{2}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{i1}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_1}{\sum_{j=1}^{d} w_j^2}\right] && (178)\\
\frac{\partial r_i^2}{\partial w_2} &= \frac{2}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{i2}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_2}{\sum_{j=1}^{d} w_j^2}\right] && (179)\\
&\vdots\\
\frac{\partial r_i^2}{\partial w_d} &= \frac{2}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{id}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_d}{\sum_{j=1}^{d} w_j^2}\right] && (180)\\
& && (181)
\end{aligned}
$$

The gradient descent updates would be

$$
\begin{aligned}
w_0^{t+1} &= w_0^t - \frac{2\alpha}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)\right] && (182)\\
w_1^{t+1} &= w_1^t - \frac{2\alpha}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{i1}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_1}{\sum_{j=1}^{d} w_j^2}\right] && (183)\\
w_2^{t+1} &= w_2^t - \frac{2\alpha}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{i2}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_2}{\sum_{j=1}^{d} w_j^2}\right] && (184)\\
&\vdots\\
w_d^{t+1} &= w_d^t - \frac{2\alpha}{\sum_{j=1}^{d} w_j^2}\left[(w_0 + \sum_{i=1}^{n}(\sum_{j=1}^{d}(w_j x_{ij} - y_i))x_{id}) - \frac{(w_0 + \sum_{i=1}^{n}\sum_{j=1}^{d} w_j x_{ij} - y_i)^2 w_d}{\sum_{j=1}^{d} w_j^2}\right] && (185)\\
& && (186)
\end{aligned}
$$

3.3 c

R^2 values for 5 Datasets

1. R squared for SSE 0.9986367866244276
   R squared for Orthogonal 0.9946995820865407
   R squared for ML 0.998636786764999

2. R squared for SSE 0.9996013163708001
   R squared for Orthogonal 0.9927019404105092
   R squared for ML 0.9996013164106964

3. R squared for SSE 0.999580727713216
   R squared for Orthogonal 0.992614348356078
   R squared for ML 0.9995807277547767

4. R squared for SSE 0.9996392732965919
   R squared for Orthogonal 0.9925527078550526
   R squared for ML 0.9996392733348016

5. R squared for SSE 0.9995547058243224
   R squared for Orthogonal 0.9926889527938757
   R squared for ML 0.9995547058749727