1).

a). Because ~~e~~ everything is going up by $w_*^T x_i$, the mean will increase by $w_*^T x_i$. Because everything goes up by the same amount, σ stays the same

b). $\arg\max p(y_i|x_i, w) = \frac{1}{20}\prod_{i=1}^{n} \exp\left(\frac{-|y_i - w_*^T x_i|}{\sigma}\right)$

$$\ln(p(y_i|x_i w)) = \ln \prod_{i=1}^{n} \frac{1}{20} \exp\left(\frac{-|y_i - w^T x_i|}{\sigma}\right)$$

$$\min\left(-n\ln(20) - \sum_{i=1}^{n} \frac{|y_i - w^T x_i|}{\sigma}\right) = -n\ln(20) - \sum_{i=1}^{n}\left(\frac{|y_i - w^T x_i|}{\sigma}\right)$$

$$\text{MLE Estimator} = \arg\min \sum_{i=1}^{n} |y_i - w^T x_i|$$

c). MLE estimator (gaussian) $\rightarrow$ $\arg\max p(y_i|x_i w) = \frac{1}{20}\prod_{i=1}^{n}\left(\frac{-(y_i - w_*^T x)^2}{\sigma}\right)$

$$\ln(p(y_i|x_i w)) = \ln \prod_{i=1}^{N} \frac{1}{20} \exp\left(\frac{-(y_i - w_*^T x)^2}{\sigma}\right)$$

$$\min\left(-n\ln(20) - \sum_{i=1}^{N} \frac{(y_i - w_*^T x_i)^2}{\cancel{\sigma}}\right) - n\ln(20) - \sum_{i=1}^{N} \frac{(y_i - w_*^T x_i)^2}{\sigma_2}$$

$$\text{gaussian MLE estimator} : \arg\min \sum_{i=1}^{N} (y_i - w_*^T x_i)^{\cancel{2}}$$

They are different because in the guassian version, $(y_i - w_*^T x_i)$ is squared (and not abs). This is more seceptible to values being altered, which means more noise.

2). $\frac{1}{2}\sum_{i=1}^{n}(w_1 x_{i,1} + w_2 x_{i,2} - y_i)^2 = f_{OLS}(w)$

$$\frac{f_{ols}(w)}{\partial w_1} = \sum_{i=1}^{N}(w_1 x_{i,1} + w_2 x_{i,2} - y_i)\cdot x_{i,1} = \sum_{i=1}^{N}(w_1 x_{i,1}) + \sum_{i=1}^{N}(w_2 x_{i,2})\cdot x_{i,1} - \sum_{i=1}^{N} y_i x_{i,1}$$

$$\sum_{i=1}^{N} w_1 x_{i,1}\cdot x_{i,1} + \sum_{i=1}^{N} w_2 x_{i,2}\cdot x_{i,1} - \sum_{i=1}^{N} y_i x_{i,1} = 0$$

① $w_1 \sum_{i=1}^{N} x_{i,1}^2 + w_2 \sum_{i=1}^{N} x_{i,2}\cdot x_{i,1} - \sum_{i=1}^{N} y_i x_{i,1} = 0$

$$\frac{f_{OLS}(w)}{\partial w_2} = \sum_{i=1}^{N}(w_1 x_{i,1} + w_2 x_{i,2} - y_i)\cdot x_{i,2} = \sum_{i=1}^{N}(w_1 x_{i,1}) + \sum_{i=1}^{N} w_2 \cdot x_{i,2}\cdot x_{i,2} - \sum_{i=1}^{N} y_i \cdot x_{i,2}$$

② $w_1 \cdot \sum_{i=1}^{N} x_{i,1}\cdot x_{i,2} + w_2 \sum_{i=1}^{N} x_{i,2}^2 - \sum_{i=1}^{N} y_i \cdot x_i$

$\sum_{i=1}^{N} x_{i,1}^2 = a \quad \sum_{i=1}^{N} x_{i,2}\cdot x_{i,1} = b \quad \sum_{i=1}^{N} y_i x_{i,1} = c$

$\sum_{i=1}^{n} x_{i,2}^2 = d \quad \sum_{i=1}^{N} y_i \cdot x_{i,2} = e$

$w_1^* a + w_2^* b - c = 0$
$w_1^* b + w_2^* d - e = 0$
$w_1^* a + w_2^* b = c$
$w_1^* b + w_2^* d = e$

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix}\begin{bmatrix} w_1^* \\ w_2^* \end{bmatrix} = \begin{bmatrix} c \\ e \end{bmatrix} \qquad \begin{bmatrix} w_1^* \\ w_2^* \end{bmatrix} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}^{-1}\begin{bmatrix} c \\ e \end{bmatrix}$$

$$\frac{1}{ad - b^2}\begin{bmatrix} d & -b \\ -b & a \end{bmatrix}\cdot\begin{bmatrix} c \\ e \end{bmatrix} \qquad w_1^* = \frac{dc - be}{ad - b^2} \qquad w_2^* = \frac{-bc + ae}{ad - b^2}$$

$$\frac{\left(\left(\sum_{i=1}^{N} x_{i,2}^2 \cdot \sum_{i=1}^{N} y_i x_{i,1}\right) - \left(\sum_{i=1}^{N} x_{i,2}\, x_{i,1} \cdot \sum_{i=1}^{N} y_i \cdot x_{i,2}\right)\right)}{\left(\sum_{i=1}^{N} x_{i,1}^2 \cdot \sum_{i=1}^{N} x_{i,2}^2\right) - \left(\sum_{i=1}^{n} x_{i,2}\cdot x_{i,1}\right)^2} = w_1^*$$

2 a). at

$$\frac{\left(\left(-\sum_{i=1}^{N} x_{i,2} \cdot x_{i,1} \cdot \sum_{i=1}^{N} y_i \cdot x_{i,1}\right) + \left(\sum_{i=1}^{N} x_{i,1}^2 \cdot \sum_{i=1}^{N} y_i \cdot x_{i,2}\right)\right)}{\left(\sum_{i=1}^{N} x_{i,1}^2 \cdot \sum_{i=1}^{N} x_{i,2}^2\right) - \left(\sum_{i=1}^{N} x_{i,2} \cdot x_{i,1}\right)^2}$$

b). $\frac{1}{2}\sum_{i=1}^{N} (w_1 x_{i,1} - y_i)^2 = f_{ols}(w) \quad \frac{\partial f_{ols}(w)}{\partial w_1} \Rightarrow \sum_{i=1}^{N} (w_1 \cdot x_{i,1} - y_i) x_{i,1}$

$= \sum_{i=1}^{N} w_1^* \cdot x_{i,1}^2 - \sum_{i=1}^{N} y_i \cdot x_{i,1} = 0 \quad w_1^* \sum_{i=1}^{N} x_{i,1}^2 = \sum_{i=1}^{N} y_i \cdot x_{i,1}$

$w_1^* = \frac{\sum_{i=1}^{N} y_i \cdot x_{i,1}}{\sum_{i=1}^{N} x_{i,1}^2} \quad$ New $= \frac{1}{2}\sum_{i=1}^{N} (w_2 \cdot x_{i,2} - y_i)^2$

$\frac{\partial f_{ols}(w)}{\partial w_2} \Rightarrow \sum_{i=1}^{N} (w_2 \cdot x_{i,2} - y_i) \cdot x_{i,2}$

$\sum_{i=1}^{N} w_2^* x_{i,2}^2 - \sum_{i=1}^{N} y_i \cdot x_{i,2} = 0 \quad w_2^* \sum_{i=1}^{N} x_{i,2}^2 = \sum_{i=1}^{N} y_i \cdot x_{i,2}$

$w_2^* = \frac{\sum_{i=1}^{N} y_i \cdot x_{i,2}}{\sum_{i=1}^{N} x_{i,2}^2} \quad w_1^* = \frac{\sum_{i=1}^{N} y_i \cdot x_{i,1}}{\sum_{i=1}^{N} x_{i,1}^2} \quad w_2^* = \frac{\sum_{i=1}^{N} y_i \cdot x_{i,2}}{\sum_{i=1}^{N} x_{i,2}^2}$

c). w_1^* vs w_2^* are different because they are looking at 2 dimensions and 1 dimension. w_1^* looks at $x_{i,1}$ and $x_{i,2}$ and that is not occuring in the 1D version. More dimensions/look at more features

3a). $\frac{1}{2}\sum_{i=1}^{N}(w_1 x_i - y_i)^2 + \lambda\|w_1\|^2 = f_{RR}(w) \Rightarrow \frac{\partial f_{RR}(w)}{\partial w_1}$

$\sum_{i=1}^{N}(w_1 x_i - y_i)\cdot x_i + 2\lambda w_1 \Rightarrow \sum_{i=1}^{N} w_1 \cdot x_i^2 - \sum_{i=1}^{N} x_i \cdot y_i + 2\lambda w_1 = 0$

$w_1^* \sum_{i=1}^{N} x_i^2 + 2\lambda = \sum_{i=1}^{N} x_i \cdot y_i \Rightarrow w_1^* = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{\sum_{i=1}^{N} x_i^2 + 2\lambda}$

b). $\frac{1}{2}\sum_{i=1}^{N}(w_1^{1*} x_i + w_2^{1*}\cdot x_i - y_i)^2 + \lambda(w^{1*2} + w_2^{1*2}) = f_{RR}(w)$

$\frac{\partial f_{RR}(w)}{\partial w_1} \quad \sum_{i=1}^{N}(w_1 x_i + w_2 x_i - y_i)\cdot x_i + 2\lambda w_1$

$= \sum_{i=1}^{N} w_1 x_i^2 + \sum_{i=1}^{N} w_2 \cdot x_i^2 - \sum_{i=1}^{N} y_i \cdot x_i + 2\lambda w_1 = 0$

$w_1\left(\sum_{i=1}^{N} x_i^2 + 2\lambda\right) + w_2 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i \cdot x_i \qquad \sum_{i=1}^{N} x_i^2 + 2\lambda = a$

$w_1^{1*}\sum_{i=1}^{N} x_i^2 + w_2^{1*}\left(\sum_{i=1}^{N} x_i^2 + 2\lambda\right) = \sum_{i=1}^{N} y_i \cdot x_i \qquad \sum_{i=1}^{N} x_i^2 = b$

$\sum_{i=1}^{N} y_i \cdot x_i = c$

$w_1^{1*} a + w_2^{1*} b = c$
$w_1^{1*} b + w_2^{1*} a = c$
$\begin{bmatrix} a & b \\ b & a \end{bmatrix}\begin{bmatrix} w_1^{1*} \\ w_2^{1*} \end{bmatrix} = \begin{bmatrix} c \\ c \end{bmatrix}$

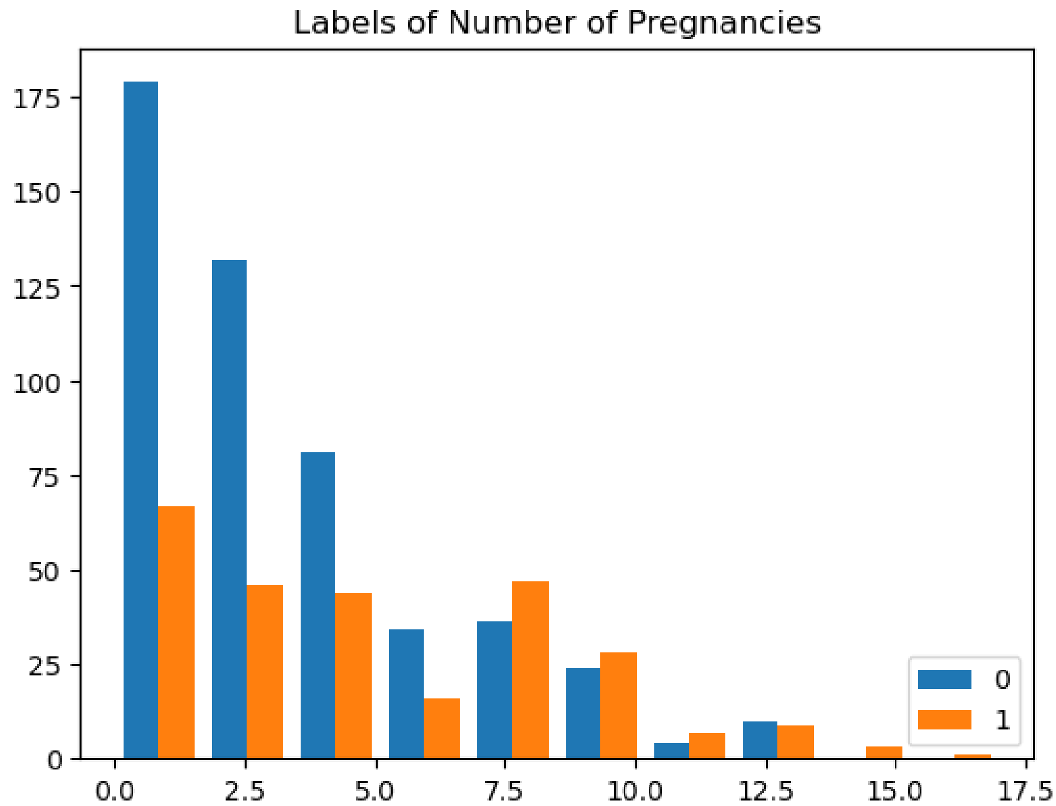
$\begin{bmatrix} w_1^{1*} \\ w_2^{1*} \end{bmatrix} = \frac{1}{a^2 - b^2}\begin{bmatrix} a & -b \\ -b & a \end{bmatrix}\begin{bmatrix} c \\ c \end{bmatrix} \Rightarrow \frac{ac - bc}{a^2 - b^2} = w_1^{*1} \qquad \frac{-bc + ac}{a^2 - b^2} = w_2^{*1} = w_1^{1*}$

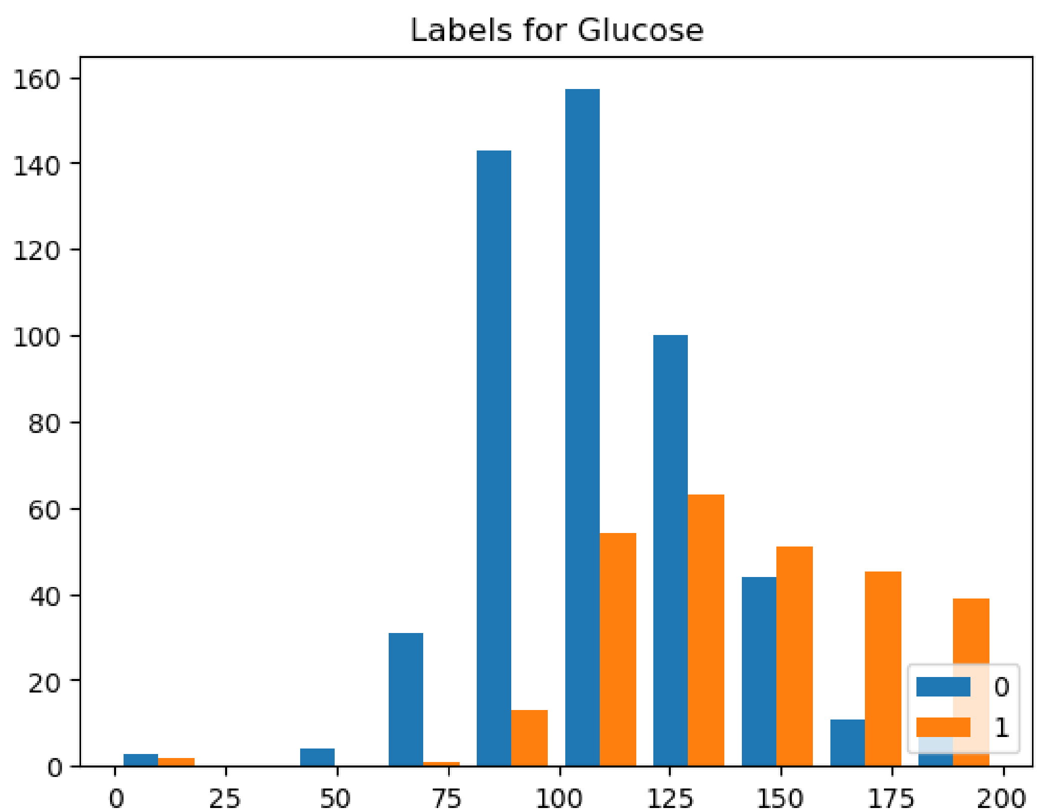
$\frac{ac - bc}{a^2 - b^2} \qquad \frac{\sum_{i=1}^{N} y_i \cdot x_i \left(\sum_{i=1}^{N} x_i^2 + 2\lambda - \sum_{i=1}^{N} x_i^2\right)}{\left(\left(\sum_{i=1}^{N} x_i^2 + 2\lambda\right)^2 - \left(\sum_{i=1}^{N} x_i^2\right)^2\right)}$

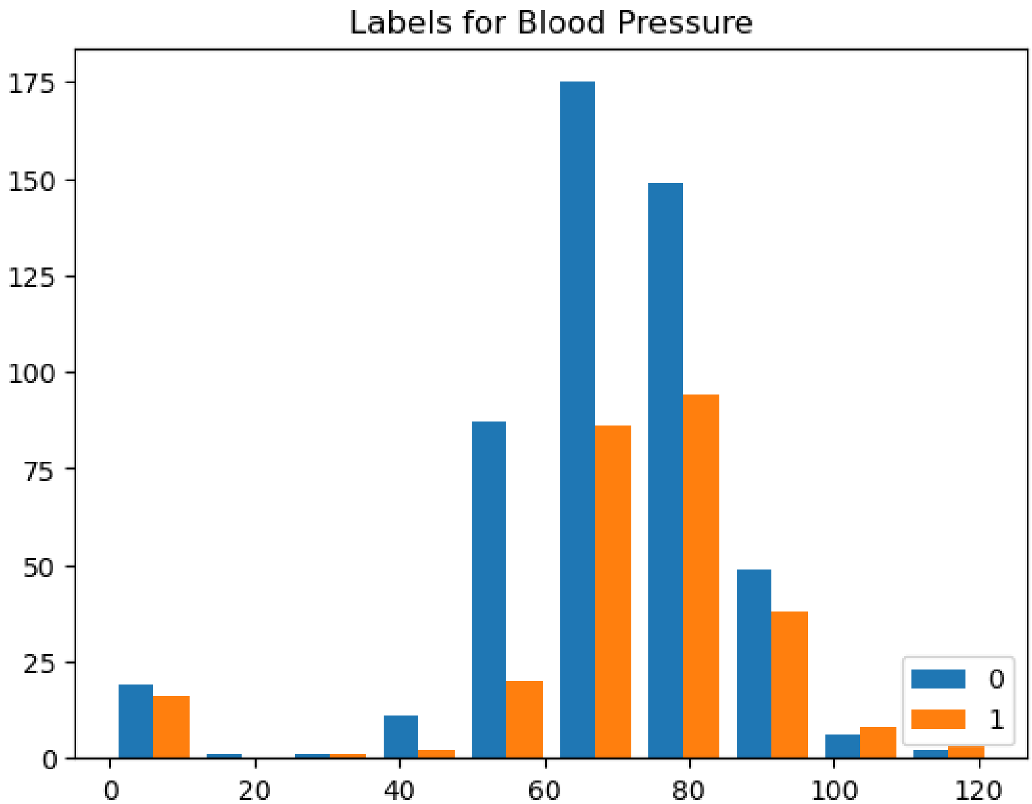
$$\frac{2\lambda \cdot \left(\sum_{i=1}^{N} y_i x_i\right)}{\left(\left(\sum_{i=1}^{N} x_i^2 + 2\lambda\right)^2 - \left(\sum_{i=1}^{N} x_i^2\right)^2\right)} = \frac{ac-bc}{a^2-b^2} = w_1^{1*} = w_2^{1*}$$

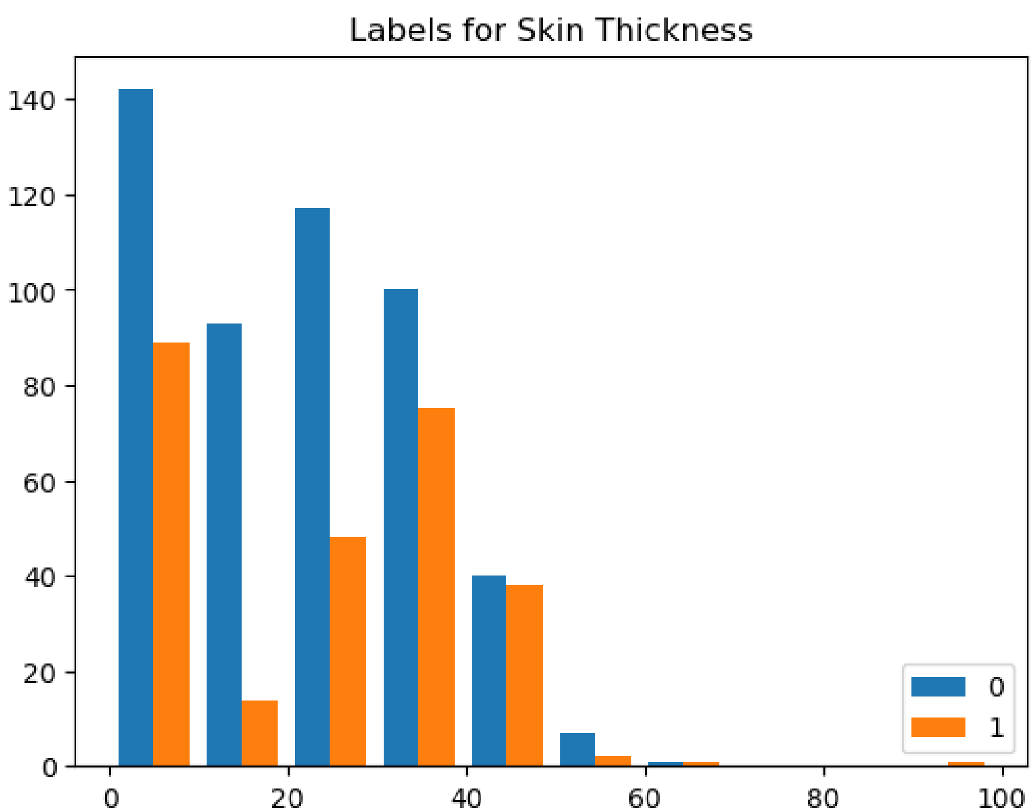
$$\frac{\sum_{i=1}^{N} y_i \cdot x_i}{2\sum_{i=1}^{N} x_i^2 + 2\lambda} = w_1^{1*} = w_2^{1*} = \frac{ac+bc}{a^2-b^2}$$

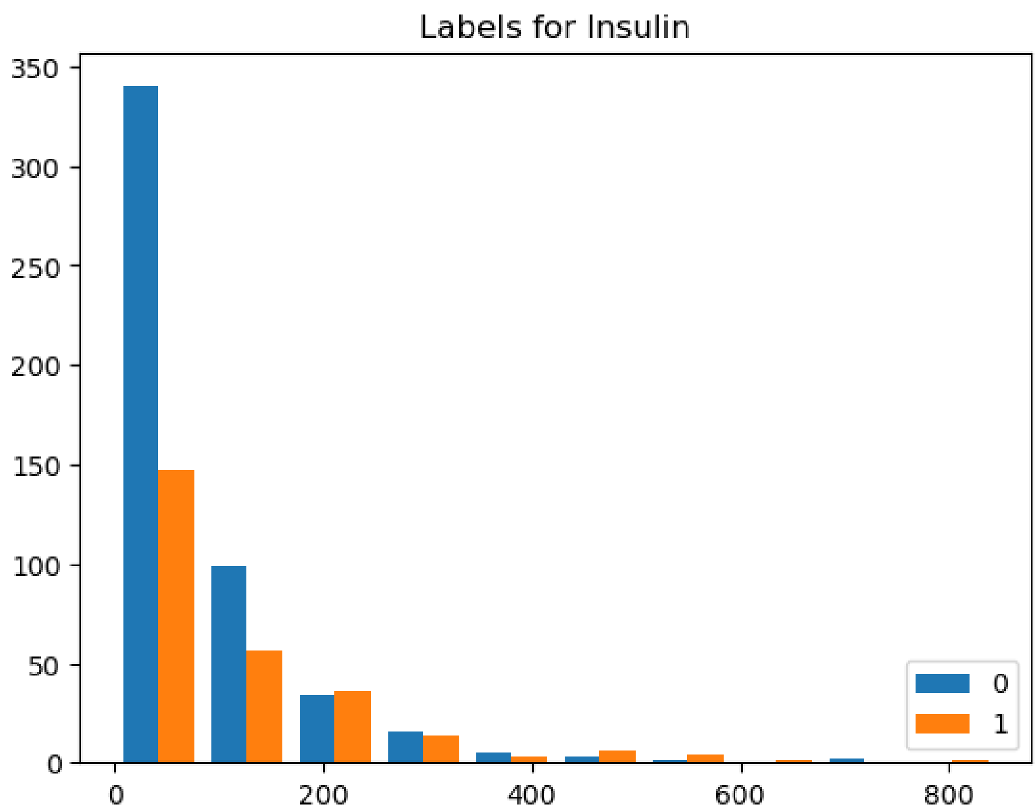
c) In w_1^{1*}, $\sum_{i=1}^{N} x_i^2$ is $\times 2$, which means w_1^{1*} is $\frac{1}{2} w_1^*$, as W_1^* weight is between w_1^{1*} and w_2^{1*}. w_1^{1*} and w_2^{1*} are $=$, so each is half of w_1^*.

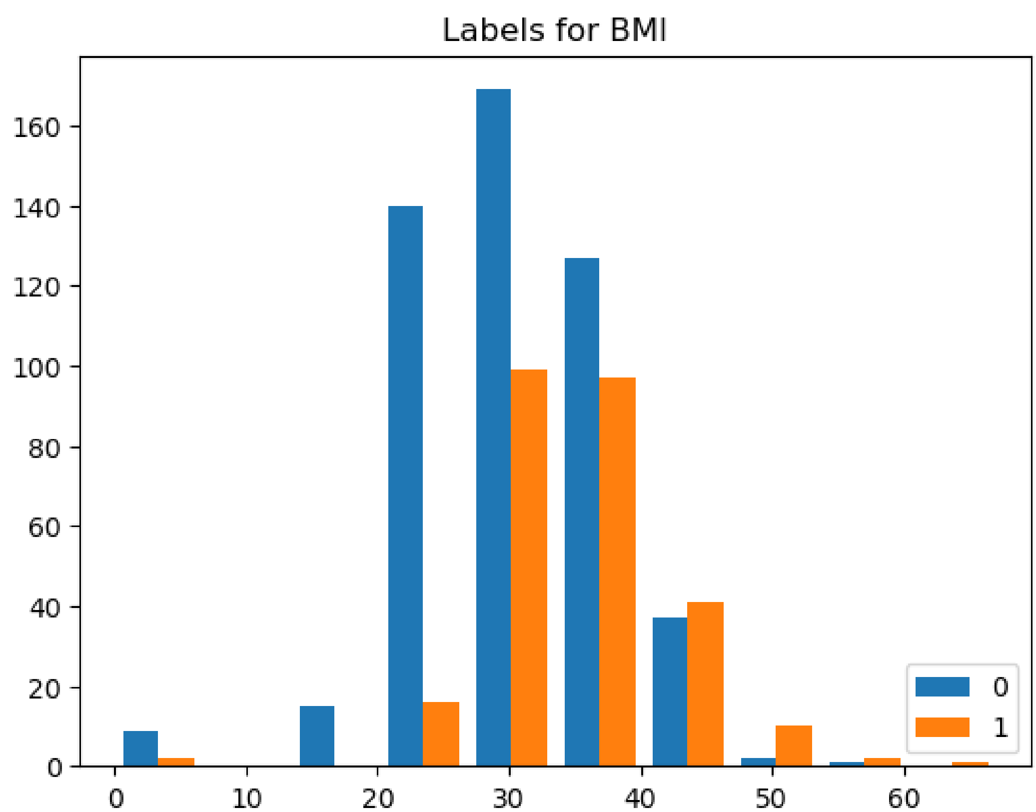

Labels of Number of Pregnancies
175
150
125
100
75
50
25
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0
1

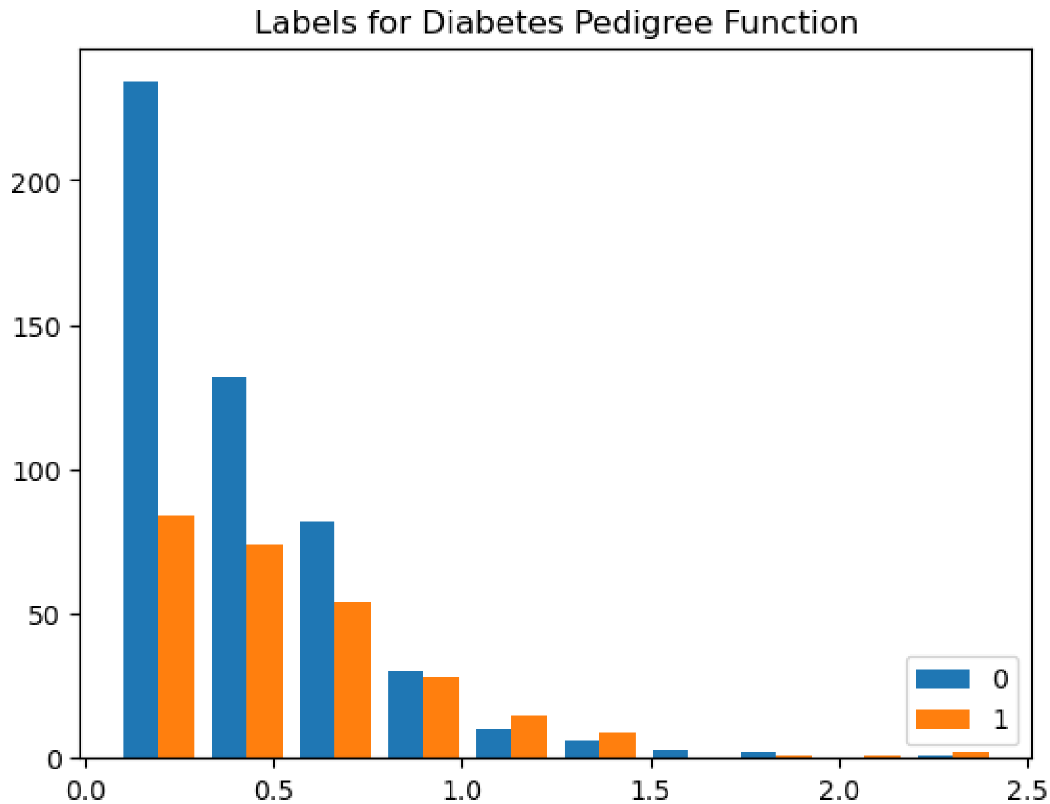

Labels for Glucose
160
140
120
100
80
60
40
20
0
0
25
50
75
100
125
150
175
200
0
1

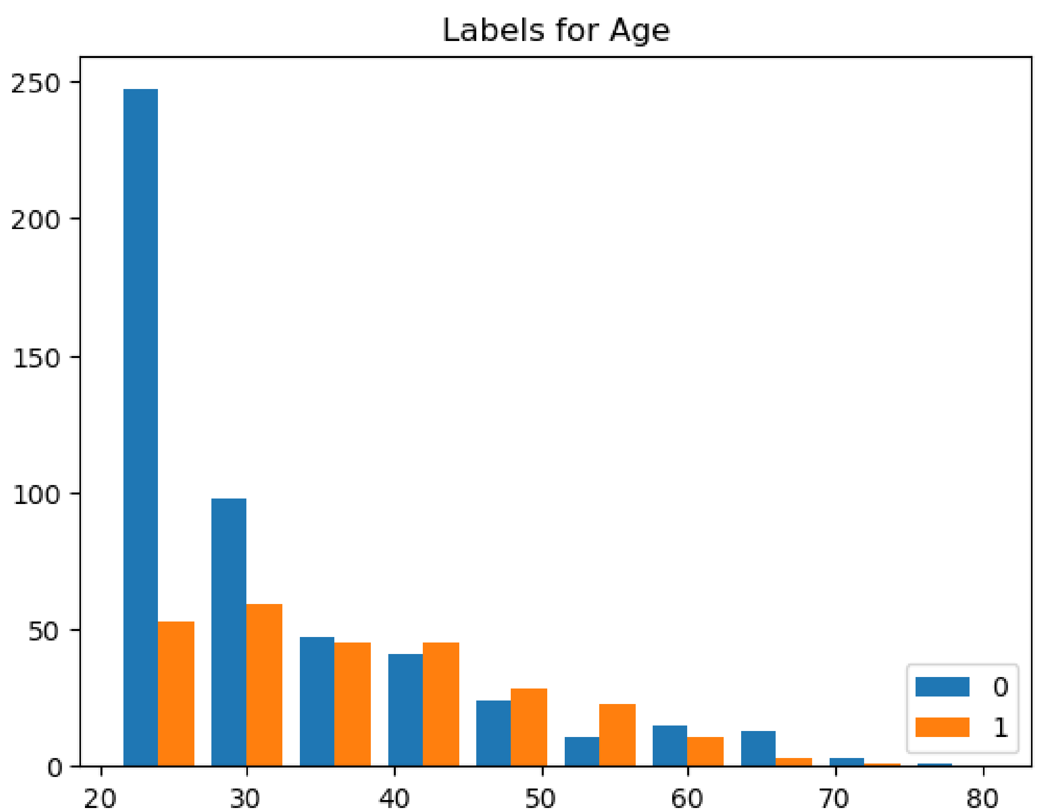

Labels for Blood Pressure
175
150
125
100
75
50
25
0
0
20
40
60
80
100
120
0
1

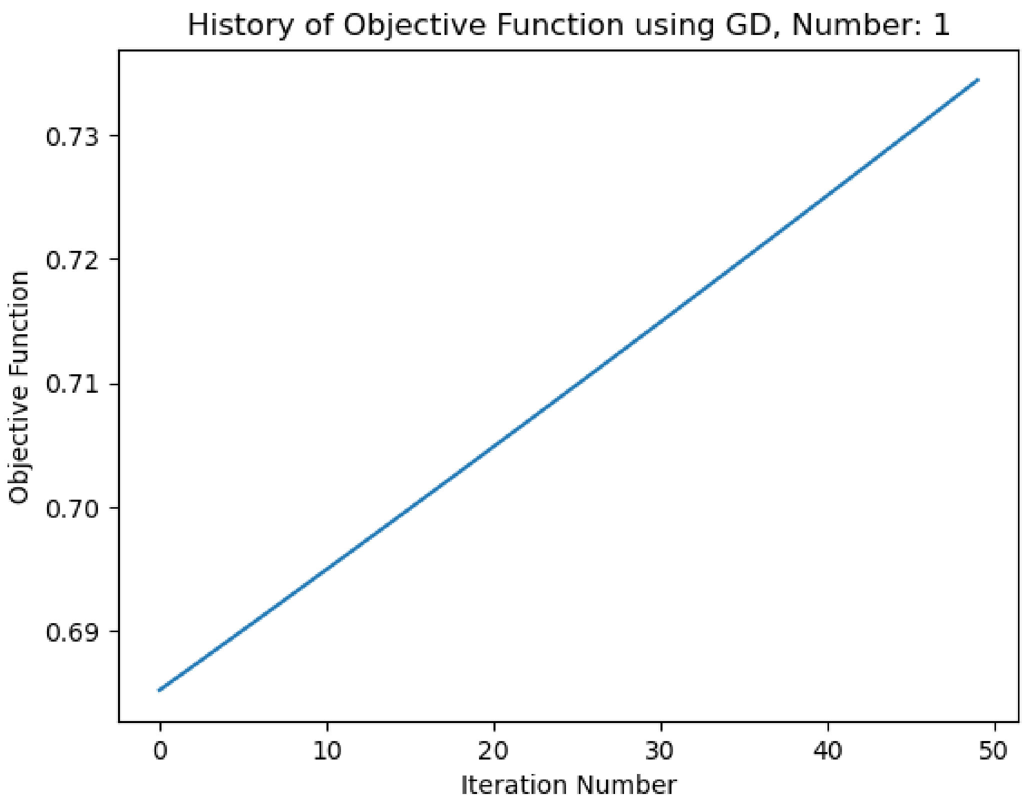

Labels for Skin Thickness
140
120
100
80
60
40
20
0
0
20
40
60
80
100
0
1

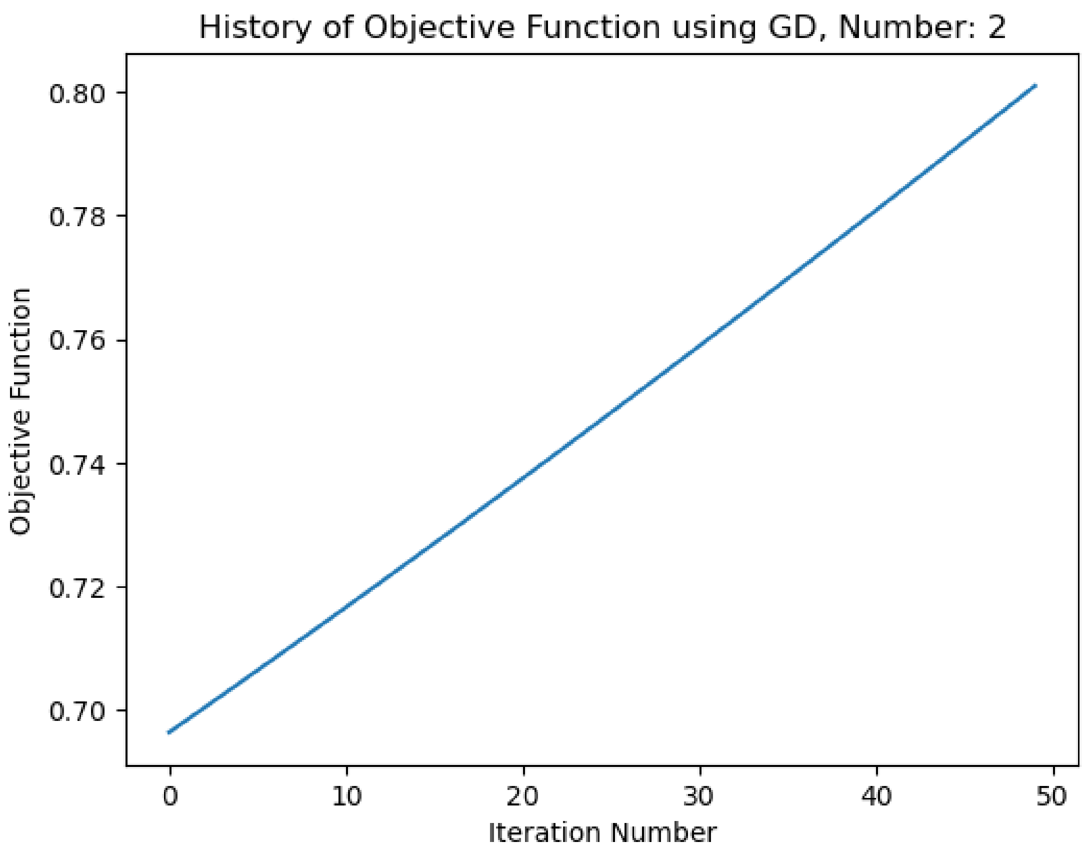

Labels for Insulin
350
300
250
200
150
100
50
0
0
200
400
600
800
0
1

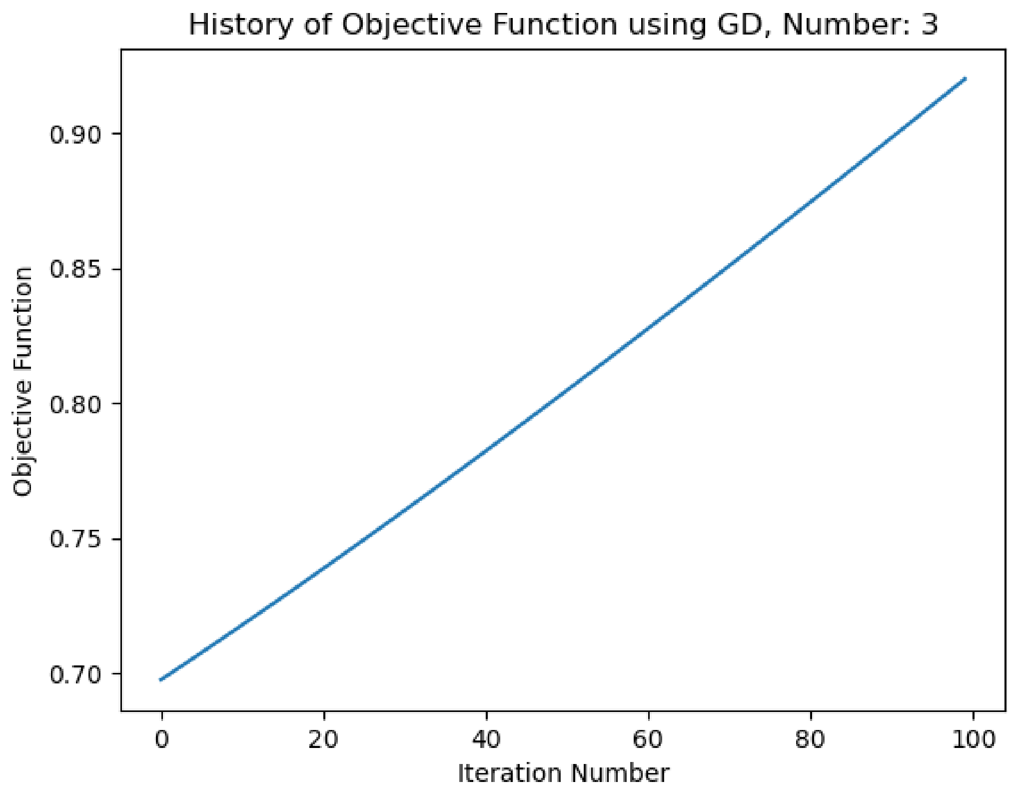

Labels for BMI
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
0
1

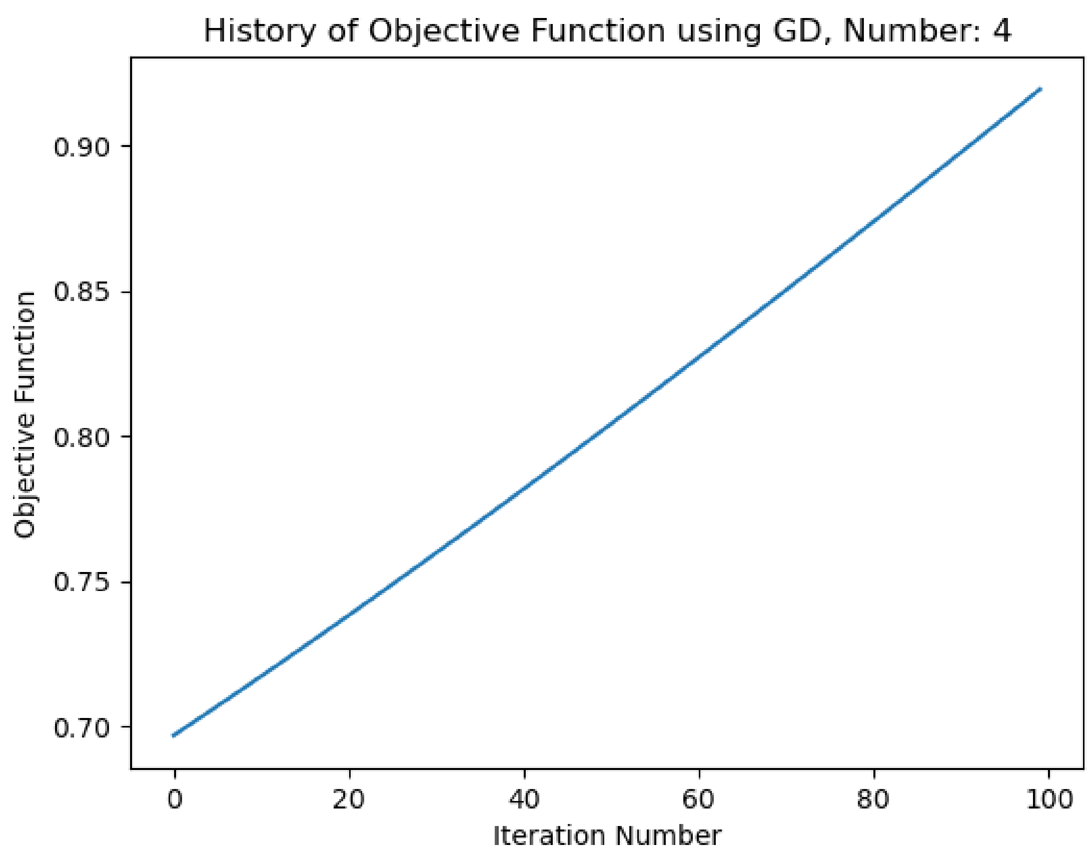

Labels for Diabetes Pedigree Function
0
50
100
150
200
0.0
0.5
1.0
1.5
2.0
2.5
0
1

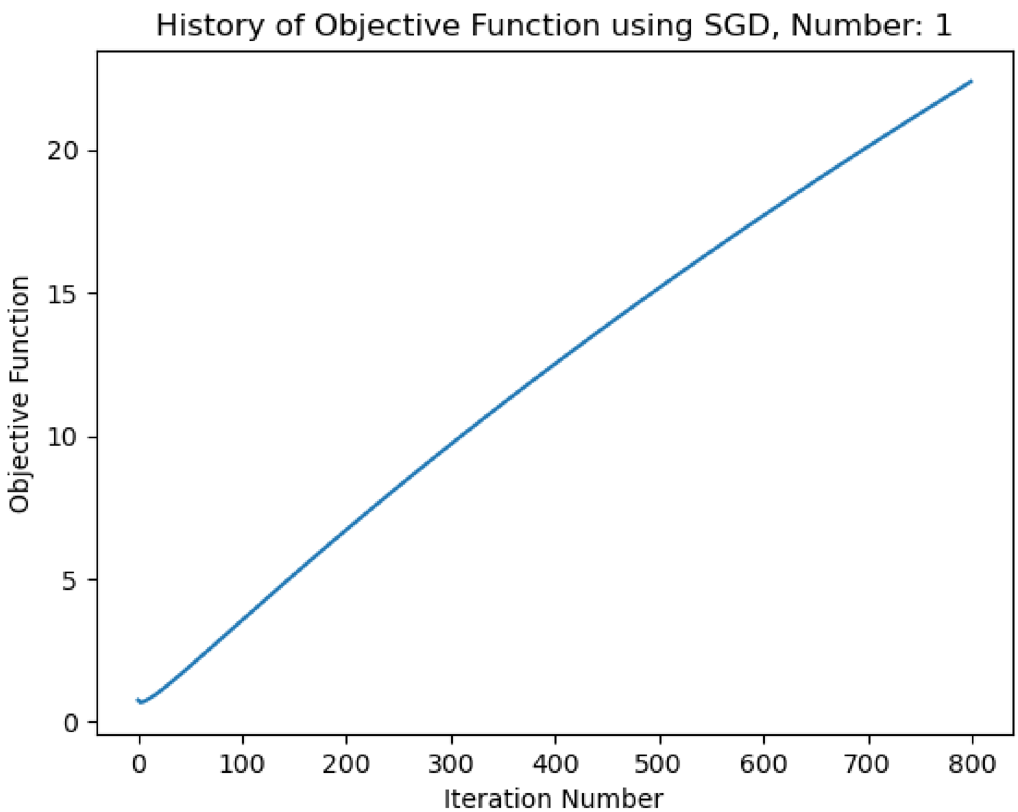

Labels for Age
0
50
100
150
200
250
20
30
40
50
60
70
80
0
1

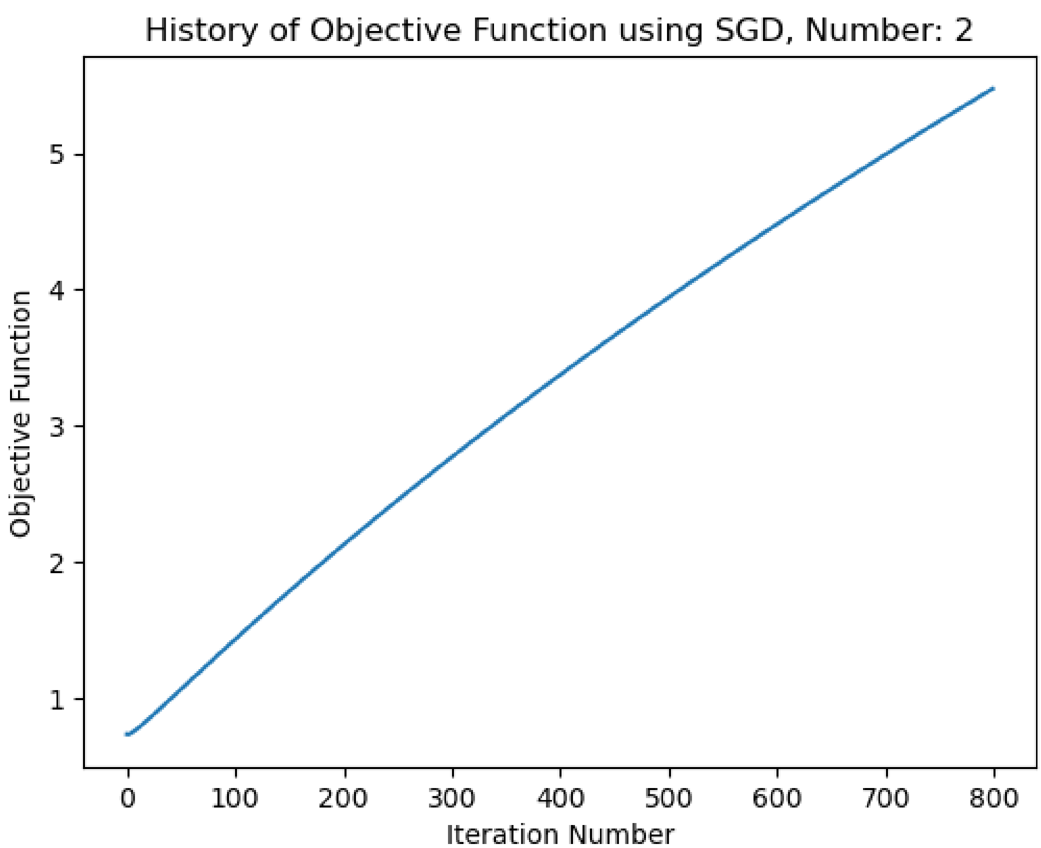

History of Objective Function using GD, Number: 1
Objective Function
Iteration Number
0.69
0.70
0.71
0.72
0.73
0
10
20
30
40
50

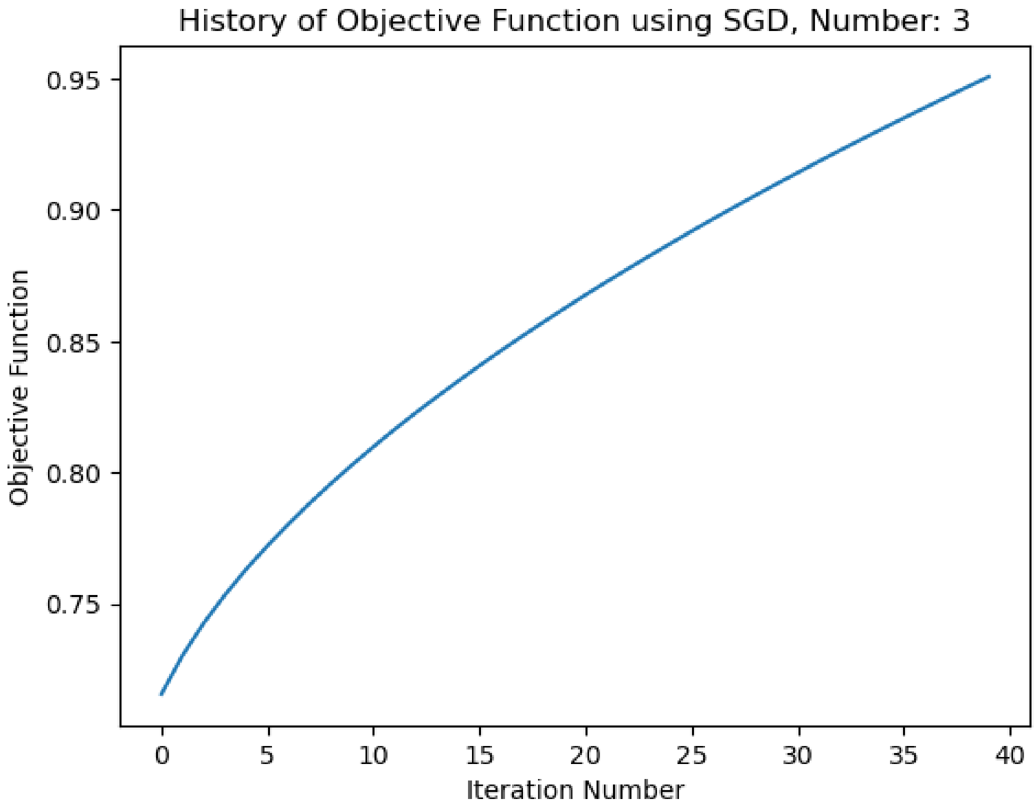

History of Objective Function using GD, Number: 2
Objective Function
Iteration Number
0.70
0.72
0.74
0.76
0.78
0.80
0
10
20
30
40
50

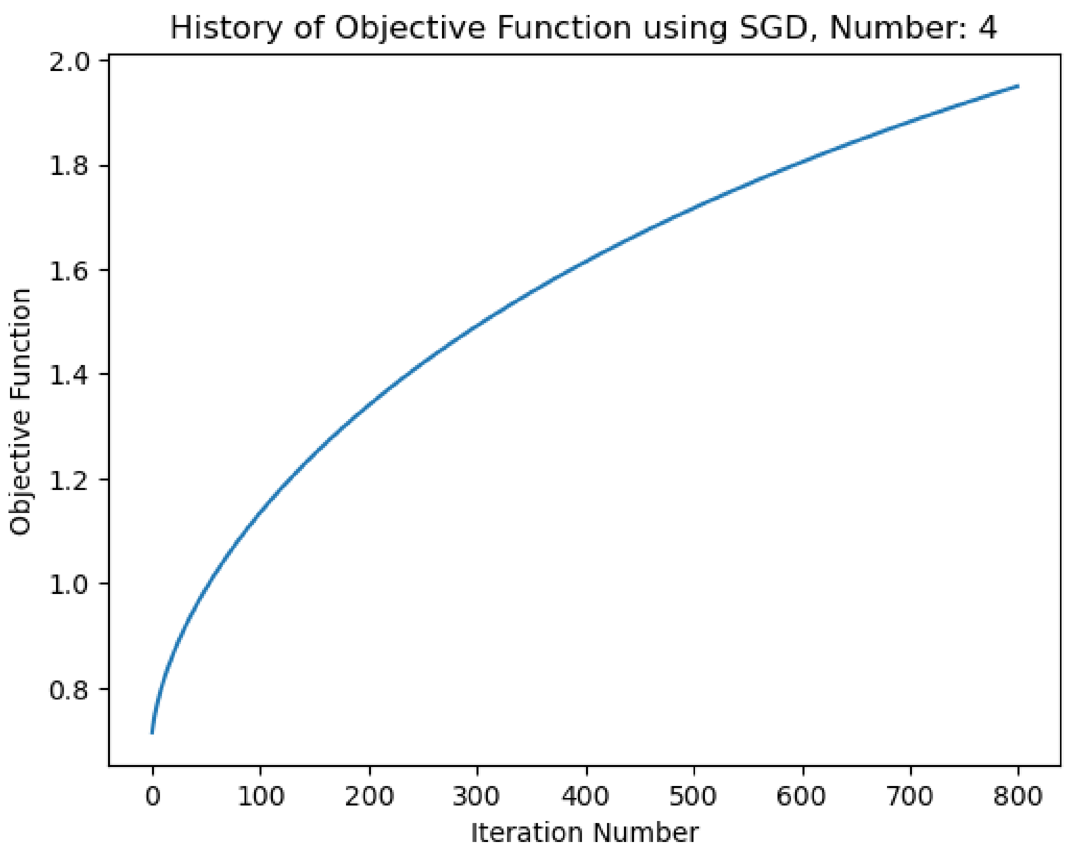

History of Objective Function using GD, Number: 3
Objective Function
0.70
0.75
0.80
0.85
0.90
0
20
40
60
80
100
Iteration Number


History of Objective Function using GD, Number: 4
Objective Function
0.70
0.75
0.80
0.85
0.90
0
20
40
60
80
100
Iteration Number


History of Objective Function using SGD, Number: 1
Objective Function
Iteration Number
0
5
10
15
20
0
100
200
300
400
500
600
700
800


History of Objective Function using SGD, Number: 2
Objective Function
Iteration Number
1
2
3
4
5
0
100
200
300
400
500
600
700
800


History of Objective Function using SGD, Number: 3
Objective Function
Iteration Number
0.95
0.90
0.85
0.80
0.75
0 5 10 15 20 25 30 35 40


History of Objective Function using SGD, Number: 4
Objective Function
Iteration Number
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0 100 200 300 400 500 600 700 800

4A: (included)

4B: The k used leads to minimization of error

4C: test error: .312

4D: Test Error: .006, here centralization and standardization helped immensely to decrease the error.