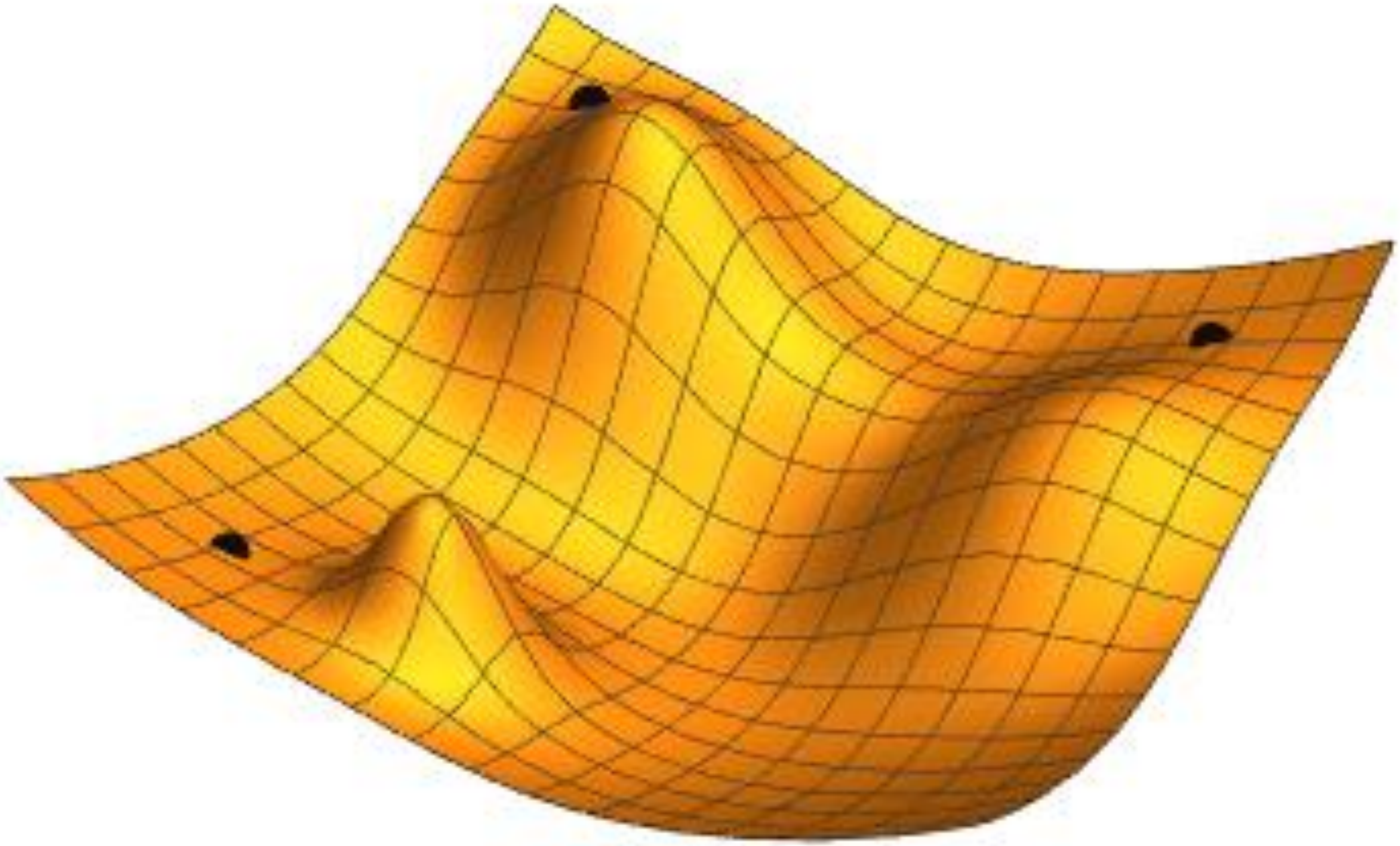
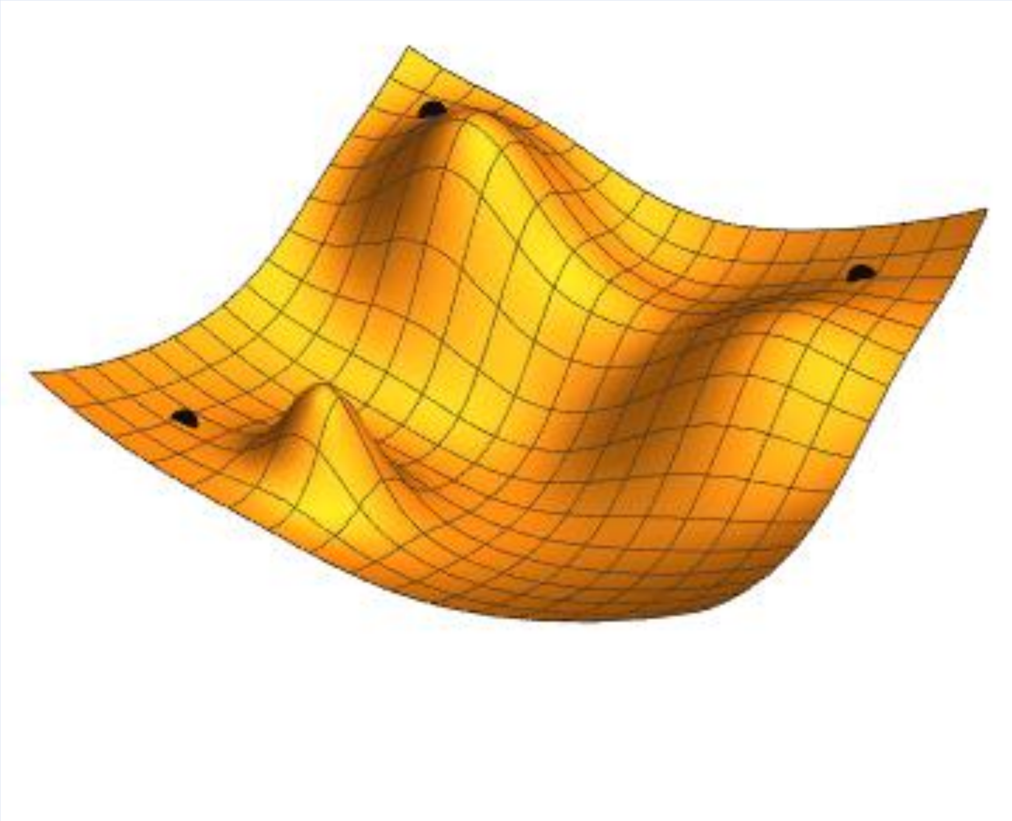


Gradient Descent

Gradient Descent

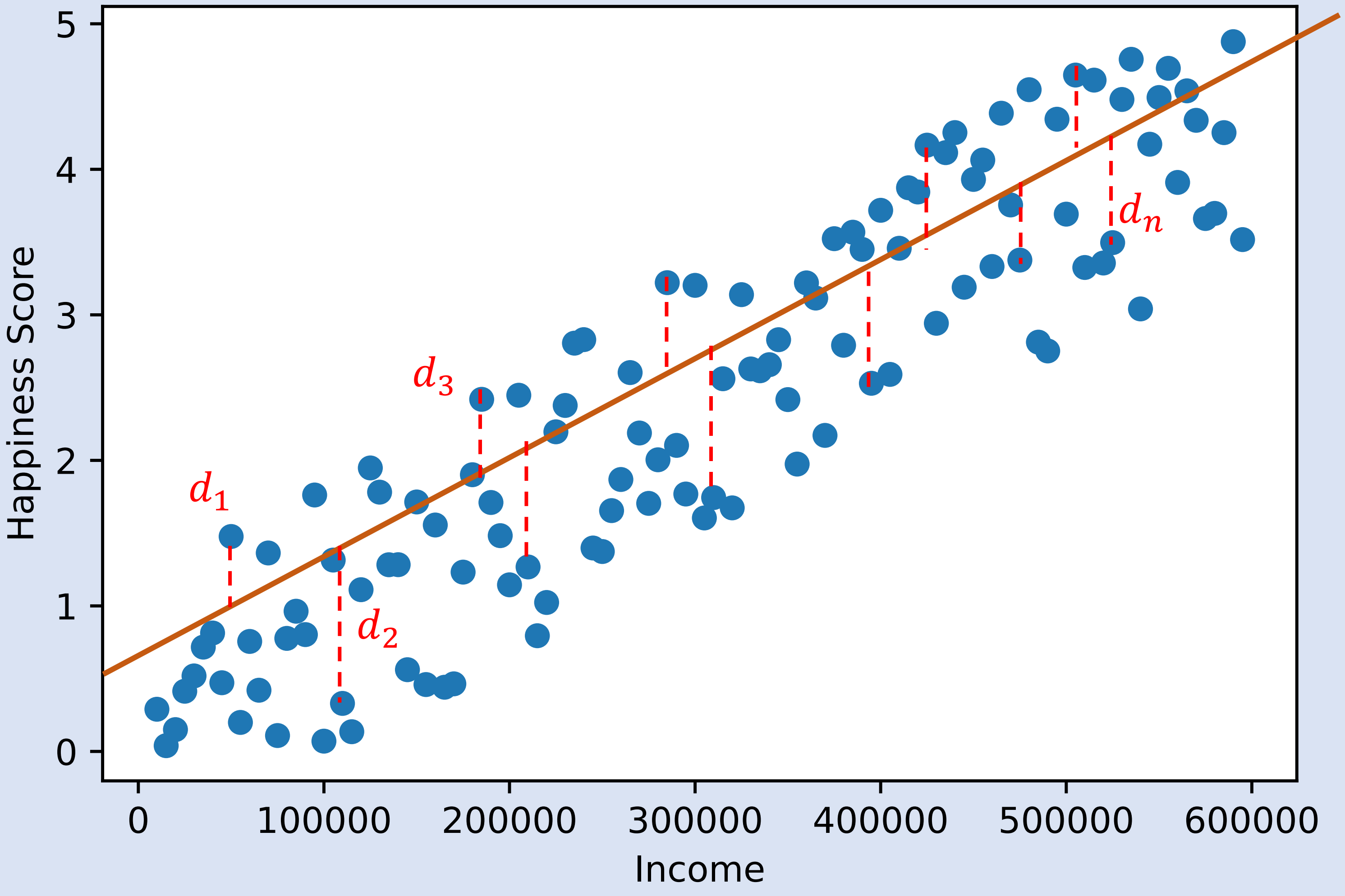
Gradient descent (also often called steepest descent) is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function



Types of Gradient Descent

- Batch Gradient Descent
- Stochastic Gradient Descent (SGD)
- Mini Batch Gradient Descent (MBGD)

Gradient Descent



$$Error = d_1 + d_2 + d_3 + \cdots + d_n$$

$$Error = d_1^2 + d_2^2 + d_3^2 + \cdots + d_n^2$$

$$Error = \sum_{i=1}^{n} d_i^2$$

$$find\ m\ and\ b\ so\ that\ the\ Error\ term\ is\ minimum$$

$$Error = \sum_{i=1}^{n} d_i^2$$

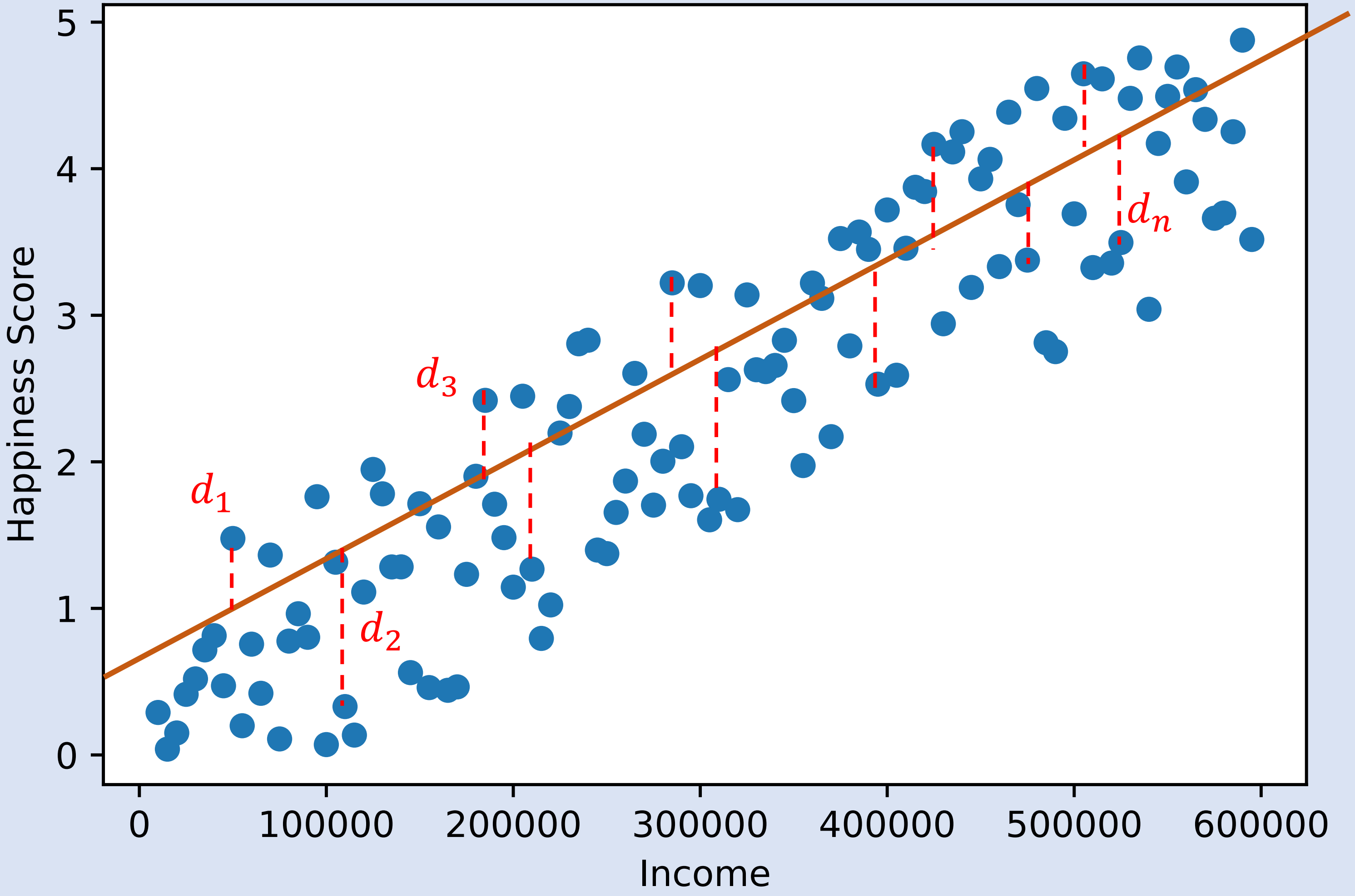
$$d_i = (y_i - \hat{y}_i)$$

$$Error = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$Sum\ of\ Squared\ Error = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Gradient Descent



$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

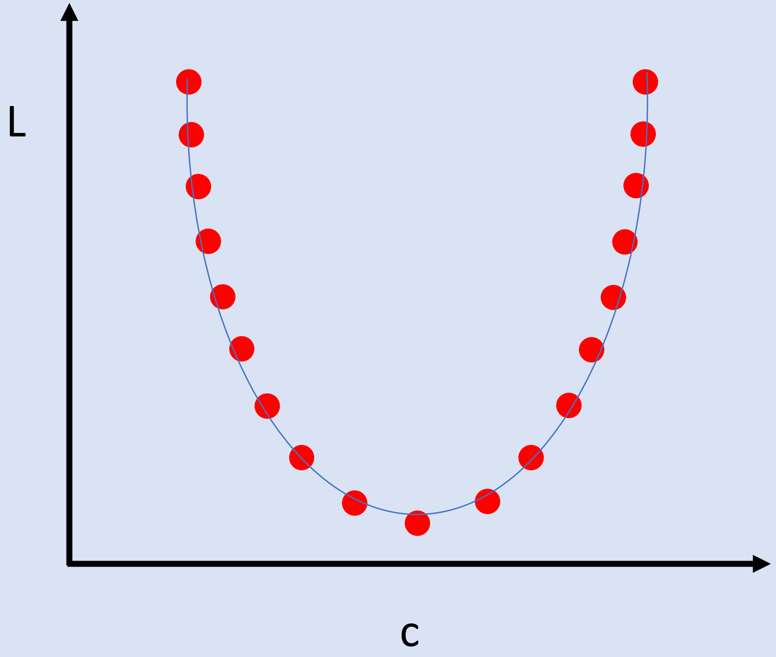
Loss Function

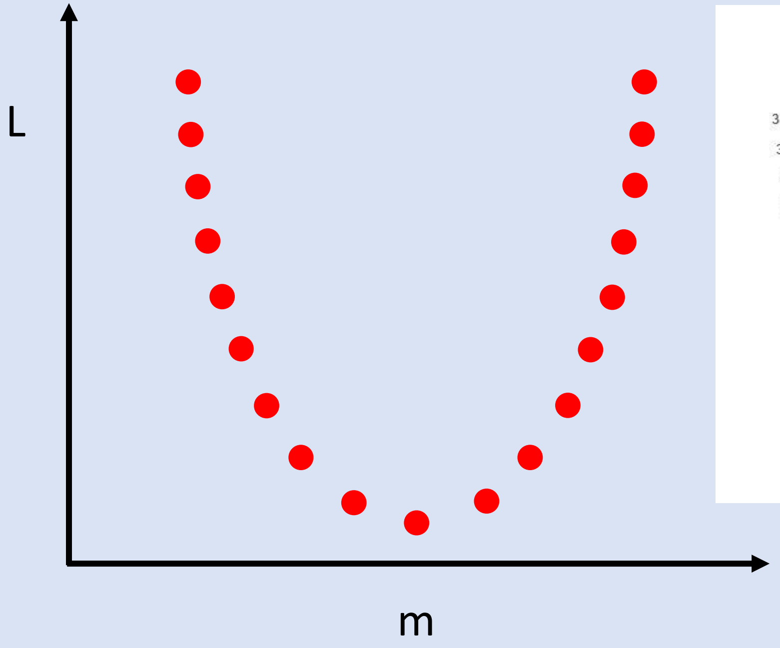
$$\hat{y}_i = mx_i + c$$

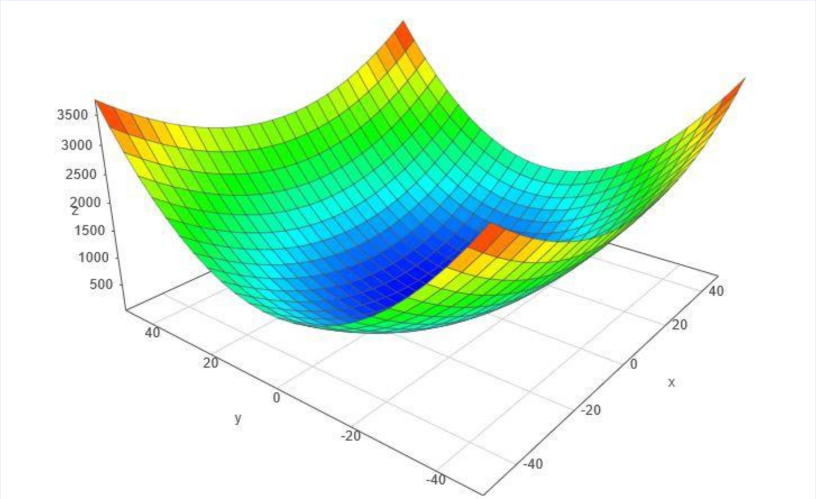
$$E(m, c) = \sum_{i=1}^{n} (y_i - mx_i - c)^2$$

$$L(m, c) = \sum_{i=1}^{n} (y_i - mx_i - c)^2$$

Gradient Descent

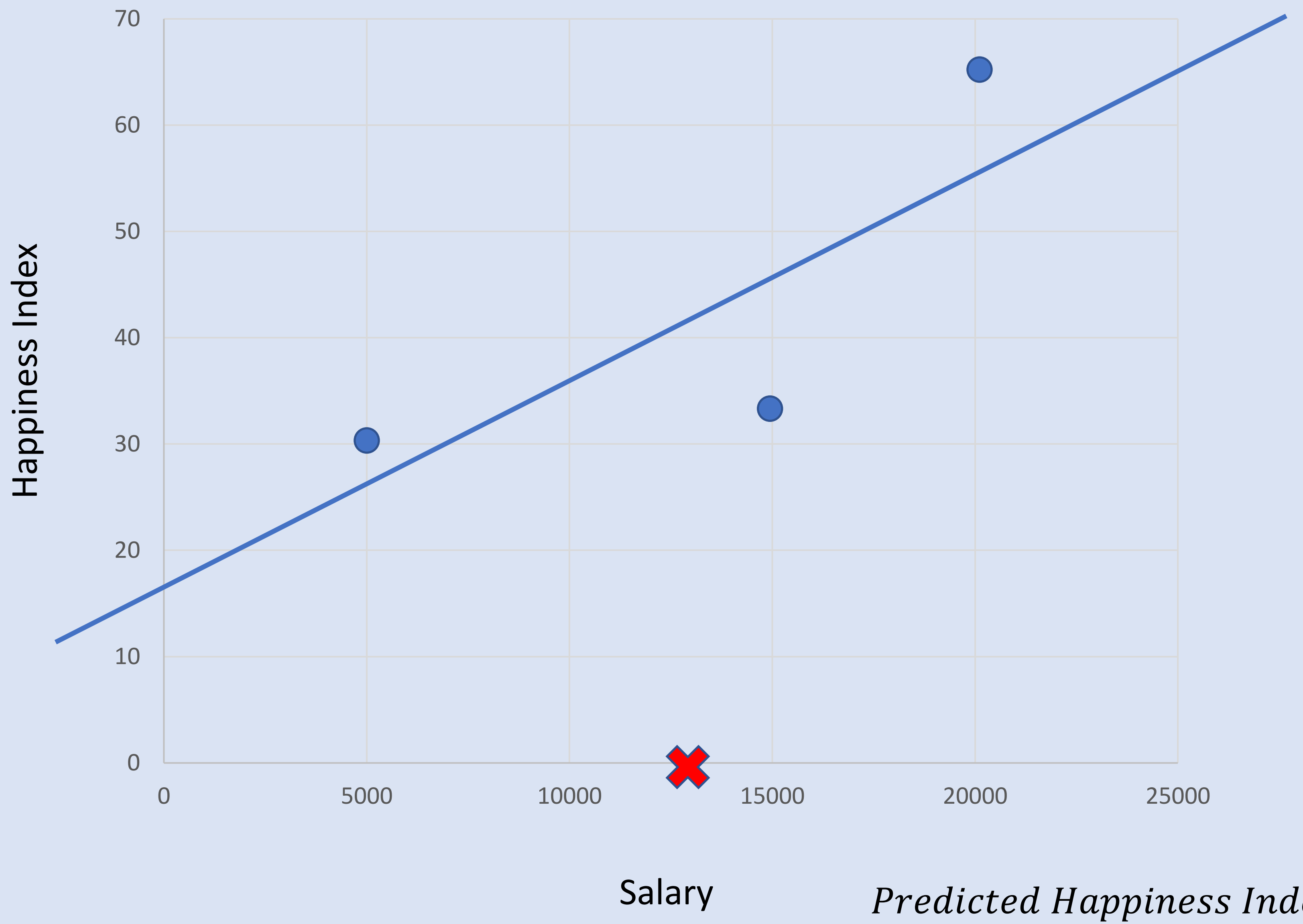






$$L(m, c) = \sum_{i=1}^{n} (y_i - mx_i - c)^2$$

$$Predicted\ Happiness\ Index = Intercept + Slope\ \times Salary$$



$$Predicted\ Happiness\ Index = Intercept + 0.002\ \times Salary$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times Salary$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times 5000 = 10$$

$$Residual = Actual - Predicted$$

$$Residual = 30 - 10$$

$$Residual = 20$$

$$Sum\ of\ Squared\ Residual = 20^2$$

$$Sum\ of\ Squared\ Residual = 400$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times Salary$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times 15000 = 30$$

$$Residual = Actual - Predicted$$

$$Residual = 33 - 30$$

$$Residual = 3$$

$$Sum\ of\ Squared\ Residual = 20^2 + 3^2$$

$$Sum\ of\ Squared\ Residual = 409$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times Salary$$

$$Predicted\ Happiness\ Index = 0 + 0.002\ \times 20000 = 40$$

$$Residual = Actual - Predicted$$

$$Residual = 65 - 40$$

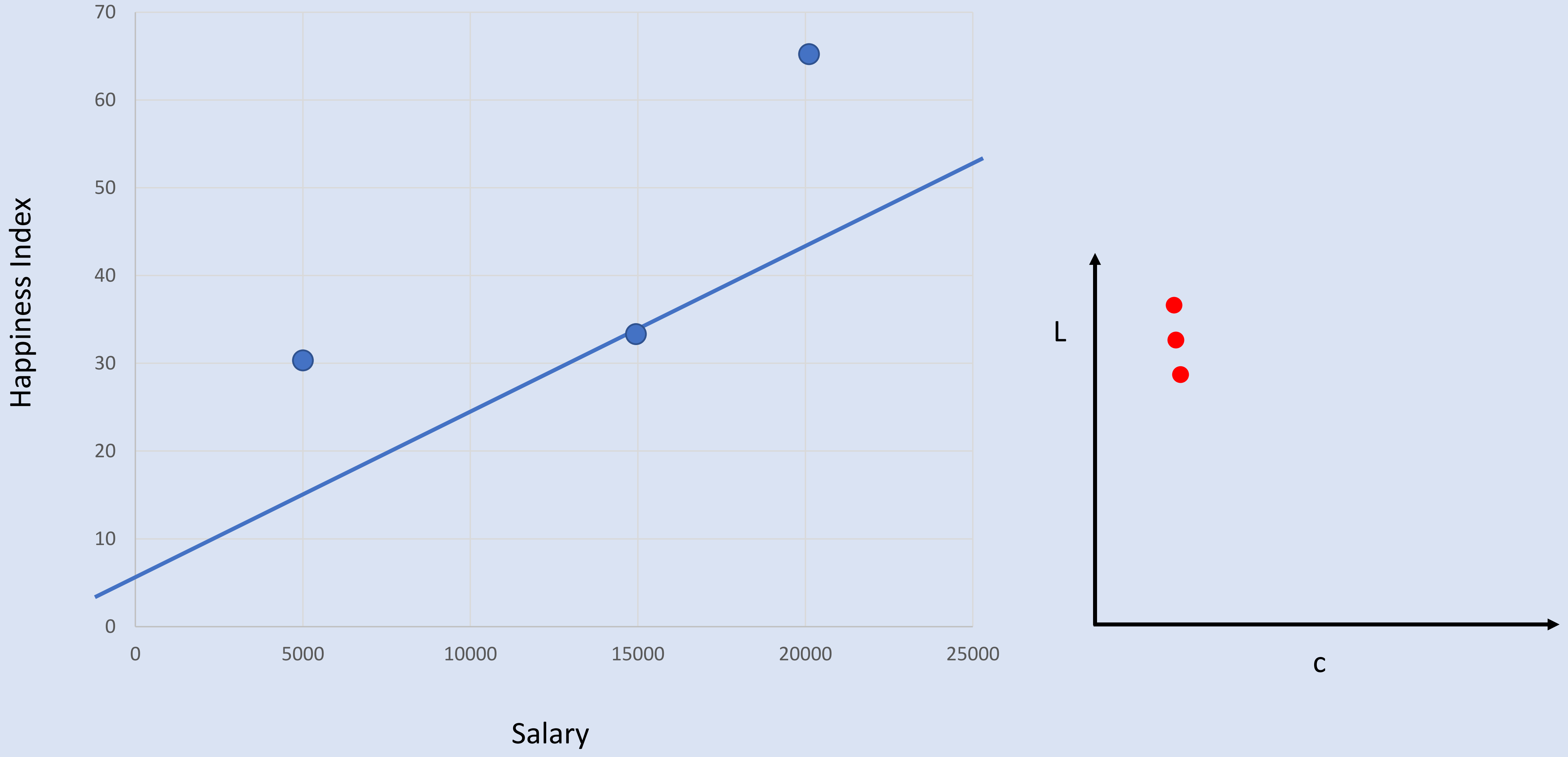
$$Residual = 15$$

$$Sum\ of\ Squared\ Residual = 20^2 + 3^2 + 15^2$$

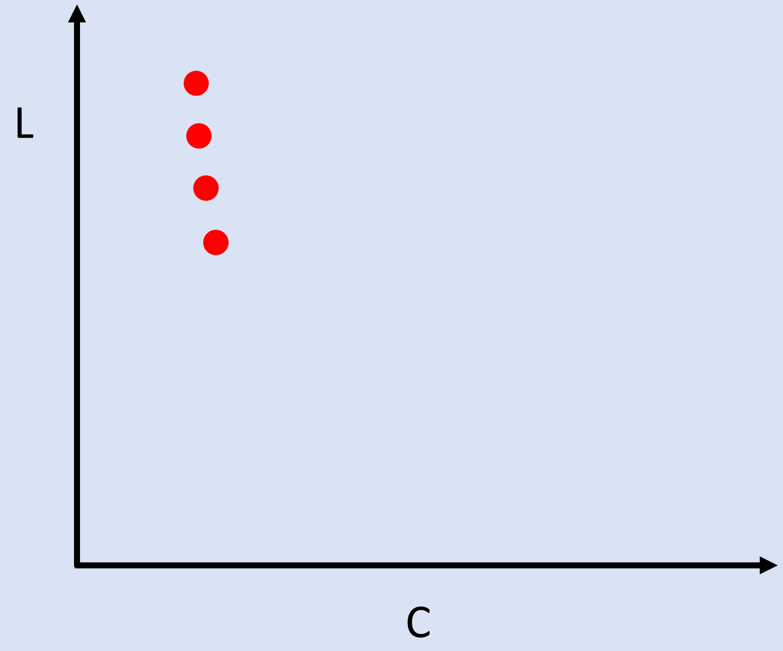
$$Sum\ of\ Squared\ Residual = 634$$

Happiness Index
70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Salary
L
c

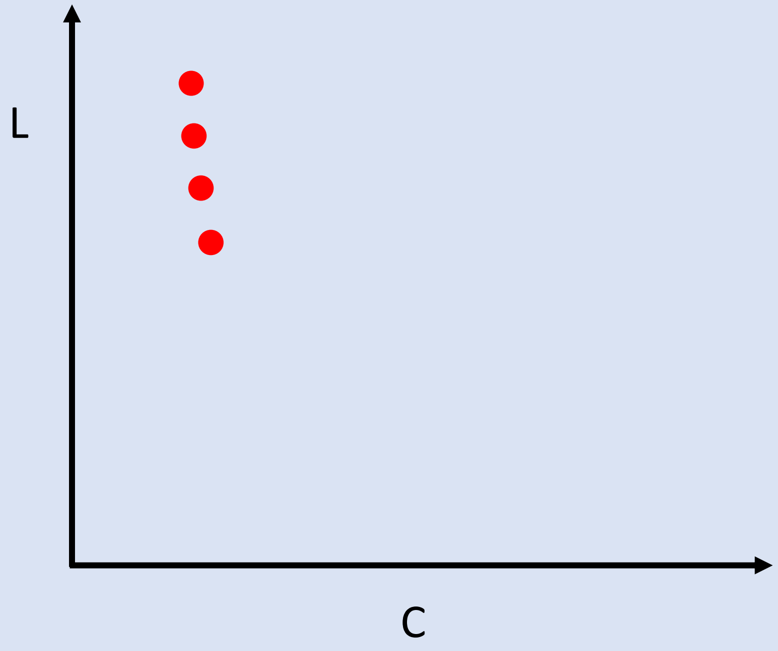
Happiness Index
70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Salary
L
c


Happiness Index
70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Salary
L
c

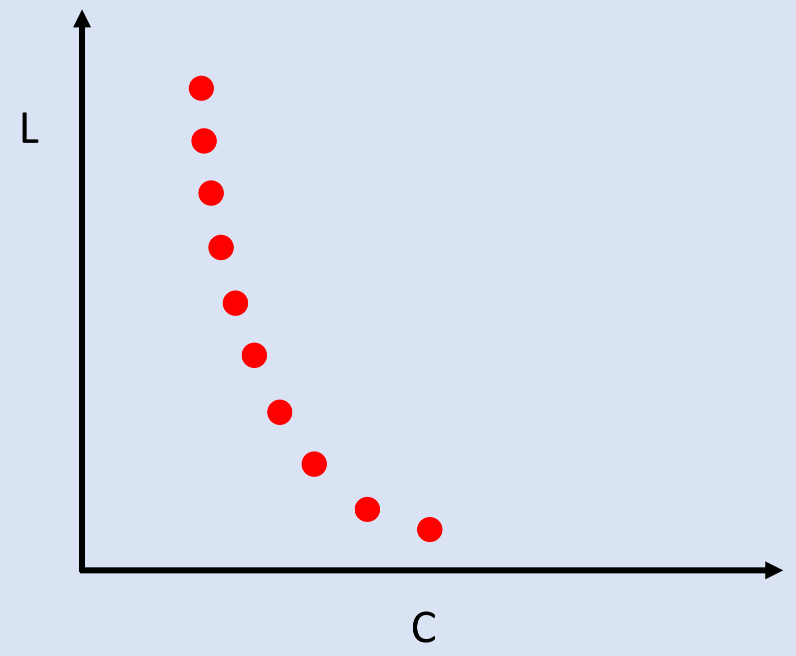

70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Happiness Index
Salary


L
C


Happiness Index
70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Salary

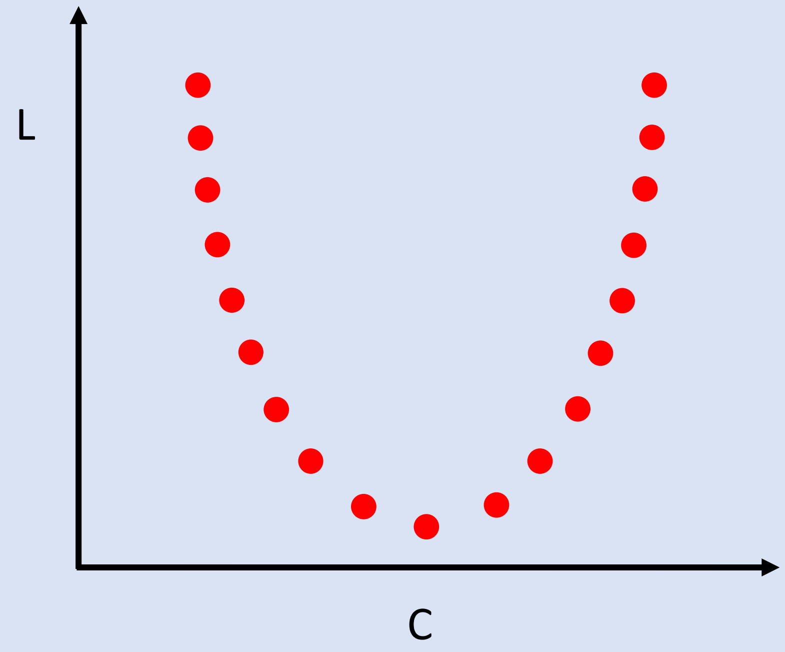

L
C

70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Happiness Index
Salary

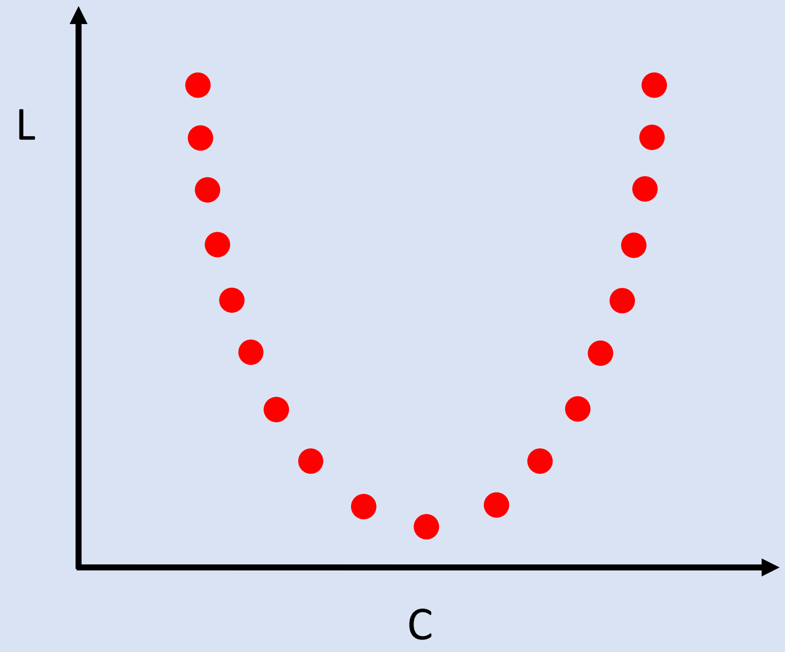


L
C


70
60
50
40
30
20
10
0
0
5000
10000
15000
20000
25000
Happiness Index
Salary


L
C

Gradient Descent Do Few Calculations



$$Sum\ of\ Squared\ Residuals = \sum (Actual - Predicted)^2$$

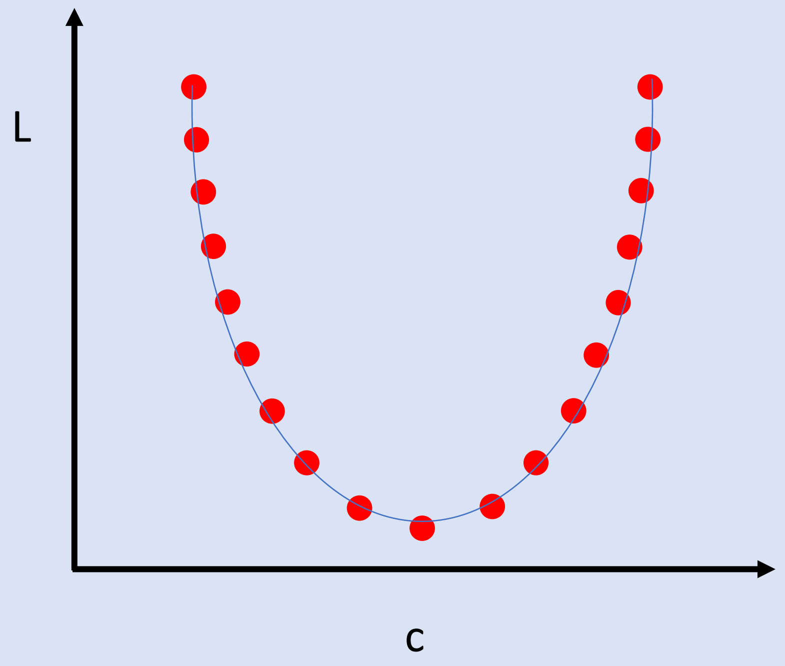
$$Sum\ of\ Squared\ Residuals = (30 - Predicted)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times Salary)\right)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times 5000)\right)^2$$

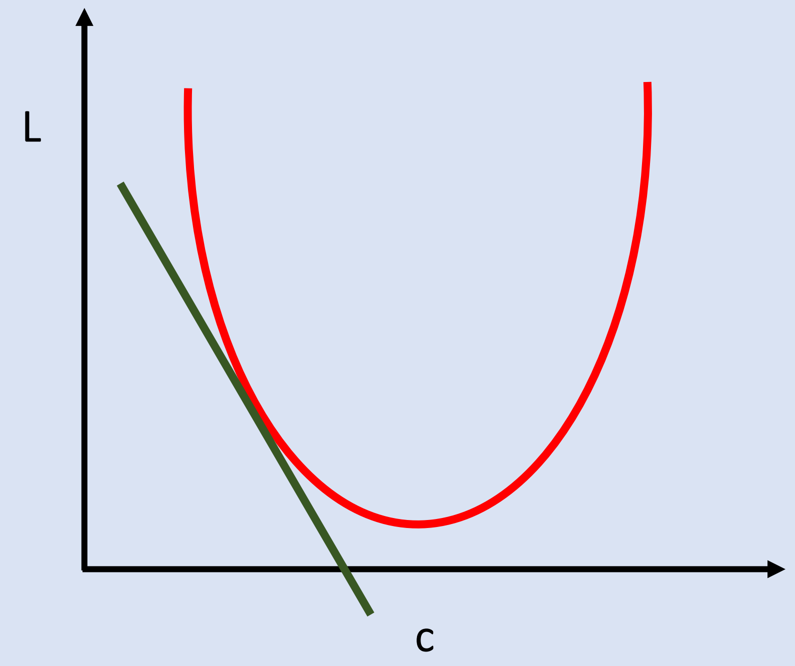
$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times Salary)\right)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times 5000)\right)^2 + \left(33 - (Intercept + 0.002 \times 15000)\right)^2 + \left(65 - (Intercept + 0.002 \times 20000)\right)^2$$



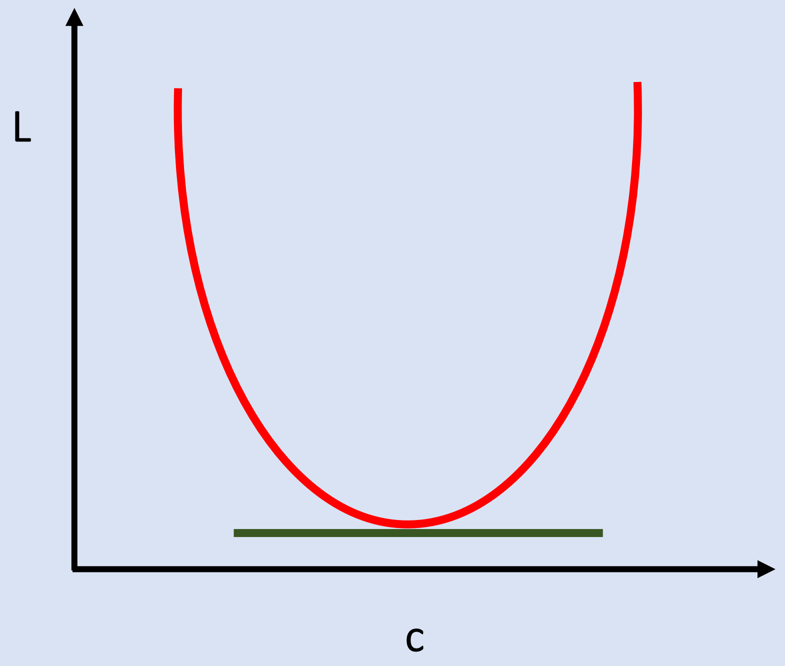
$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times Salary)\right)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times 5000)\right)^2 + \left(33 - (Intercept + 0.002 \times 15000)\right)^2 + \left(65 - (Intercept + 0.002 \times 20000)\right)^2$$



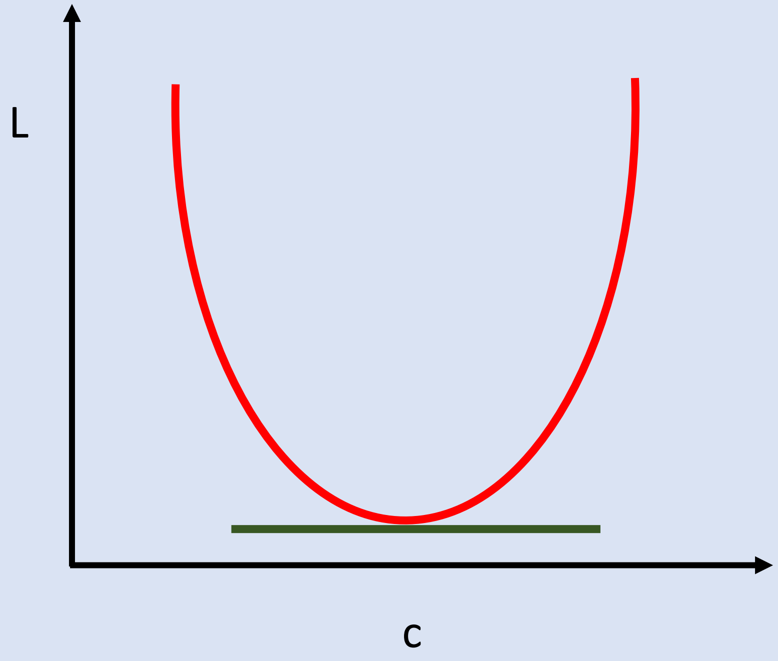
$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times Salary)\right)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times 5000)\right)^2 + \left(33 - (Intercept + 0.002 \times 15000)\right)^2 + \left(65 - (Intercept + 0.002 \times 20000)\right)^2$$



$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times Salary)\right)^2$$

$$Sum\ of\ Squared\ Residuals = \left(30 - (Intercept + 0.002 \times 5000)\right)^2 + \left(33 - (Intercept + 0.002 \times 15000)\right)^2 + \left(65 - (Intercept + 0.002 \times 20000)\right)^2$$



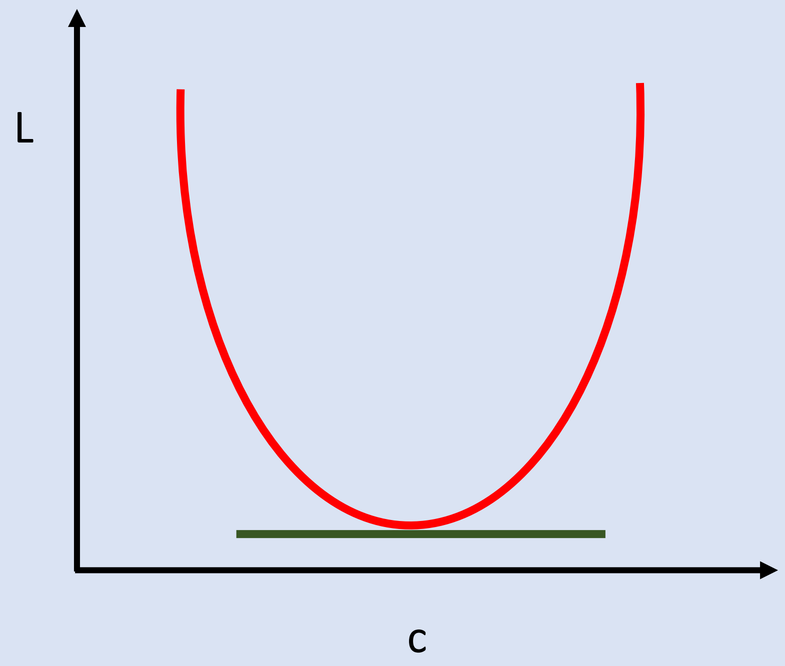
$$L = \left(30 - (C + 0.002 \times 5000)\right)^2 + \left(33 - (C + 0.002 \times 15000)\right)^2 + \left(65 - (C + 0.002 \times 20000)\right)^2$$

$$\frac{dL}{dC} = 2\left(30 - (C + 0.002 \times 5000)\right) \times (-1) + 2\left(33 - (C + 0.002 \times 15000)\right) \times (-1) + 2\left(65 - (C + 0.002 \times 20000)\right) \times (-1)$$

$$\frac{dL}{dC} = -2\left(30 - (C + 0.002 \times 5000)\right) - 2\left(33 - (C + 0.002 \times 15000)\right) - 2\left(65 - (C + 0.002 \times 20000)\right)$$

If we were to use OLS i.e. ordinary least squares, we would have simply compared $\frac{dL}{dC}$ to zero

But, Gradient Descent would compute the slope by inserting values of C in above equation step by step



$$\frac{dL}{dC} = -2(30-(C+0.002\times5000)) \\ -2(33-(C+0.002\times15000)) \\ -2(65-(C+0.002\times20000))$$

Lets compute the slope at C = 0

$$\frac{dL}{dC} = -2(30-(0+0.002\times5000)) \\ -2(33-(0+0.002\times15000)) \\ -2(65-(0+0.002\times20000))$$

$$\frac{dL}{dC} = -96$$

Learning Rate: 0.1

$$Step\ Size = -96\times0.1$$

$$Step\ Size = -9.6$$

$$\frac{dL}{dC} = -2(30-(9.6+0.002\times5000)) \\ -2(33-(9.6+0.002\times15000)) \\ -2(65-(9.6+0.002\times20000))$$

$$\frac{dL}{dC} = -38.6$$

$$Step\ Size = -38.6\times0.1$$

$$Step\ Size = -3.86$$

$$New\ Intercept = 9.6-(-3.68) \\ = 13.28$$

$$\frac{dL}{dC} = -2(30-(13.28+0.002\times5000)) \\ -2(33-(13.28+0.002\times15000)) \\ -2(65-(13.28+0.002\times20000))$$

$$\frac{dL}{dC} = -16.32$$

$$Next\ Value\ to\ put\ in\ equation\ of\ slope\ is\ calculated\ by\ substracting -9.6\ from\ initial\ Intercept$$

$$New\ Intercept\ to\ be\ tested = Old\ intercept\ - Step\ Size \ggg\gg 0-(-9.6)\gg 9.6$$

$$\frac{dL}{dC} = -2(30 - (C + 0.002 \times 5000)) \\ -2(33 - (C + 0.002 \times 15000)) \\ -2(65 - (C + 0.002 \times 20000))$$

Lets compute the slope at C = 0

$$\frac{dL}{dC} = -2(30 - (0 + 0.002 \times 5000)) \\ -2(33 - (0 + 0.002 \times 15000)) \\ -2(65 - (0 + 0.002 \times 20000))$$

$$\frac{dL}{dC} = -96$$

Learning Rate: 0.1

$$Step\ Size = -96 \times 0.1$$

$$Step\ Size = -9.6$$

$$\frac{dL}{dC} = -2(30 - (9.6 + 0.002 \times 5000)) \\ -2(33 - (9.6 + 0.002 \times 15000)) \\ -2(65 - (9.6 + 0.002 \times 20000))$$

$$\frac{dL}{dC} = -38.6$$

$$Step\ Size = -38.6 \times 0.1$$

$$Step\ Size = -3.86$$

$$New\ Intercept = 9.6 - (-3.68) \\ = 13.28$$

$$\frac{dL}{dC} = -2(30 - (13.28 + 0.002 \times 5000)) \\ -2(33 - (13.28 + 0.002 \times 15000)) \\ -2(65 - (13.28 + 0.002 \times 20000))$$

$$\frac{dL}{dC} = -16.32$$

$$Step\ Size = -16.32 \times 0.1$$

$$Step\ Size = -1.632$$

$$New\ Intercept \\ = 13.28 - (-1.632) = 14.912$$

How long it will continue?
As long as the step size comes
closer to zero
Which also corresponds to slope of
curve near to zero,
Or the number of iterations to be taken
Are manually set could be 100 or 1000
or 2000

$$Next\ Value\ to\ put\ in\ equation\ of\ slope\ is\ calculated\ by\ substracting - 9.6\ from\ initial\ Intercept$$

$$New\ Intercept\ to\ be\ tested = Old\ intercept\ - Step\ Size \ggg\gg 0 - (-9.6) \gg 9.6$$

To Sum Up….

1. Take derivative of loss function with respect to each parameters (slope and intercept) Now this derivatives taken is called as gradient

2. Later choose any random values for the parameters

3. Use these random values to compute derivatives (gradient)

4. Calculate the step sizes using formula:

Step Size = Parameter X Learning Rate

Parameter could be Slope or Intercept

5. Calculate value of new parameters to be used to compute the derivatives using formula..

New value of Parameter = Old Value of Parameter – Step Size