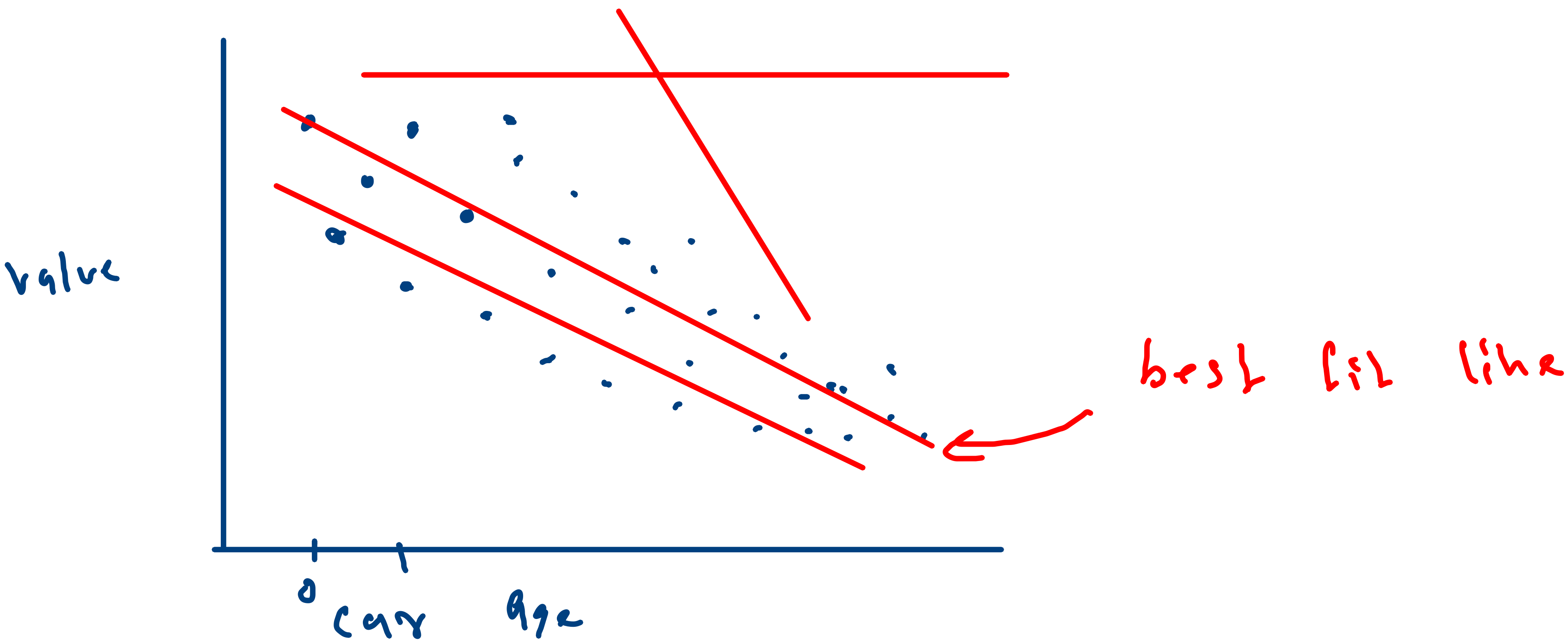
Regression → infinite set

Classification → finite set

Linear Regression

X → Y

$x_1\ x_2\ x_3\ x_n$

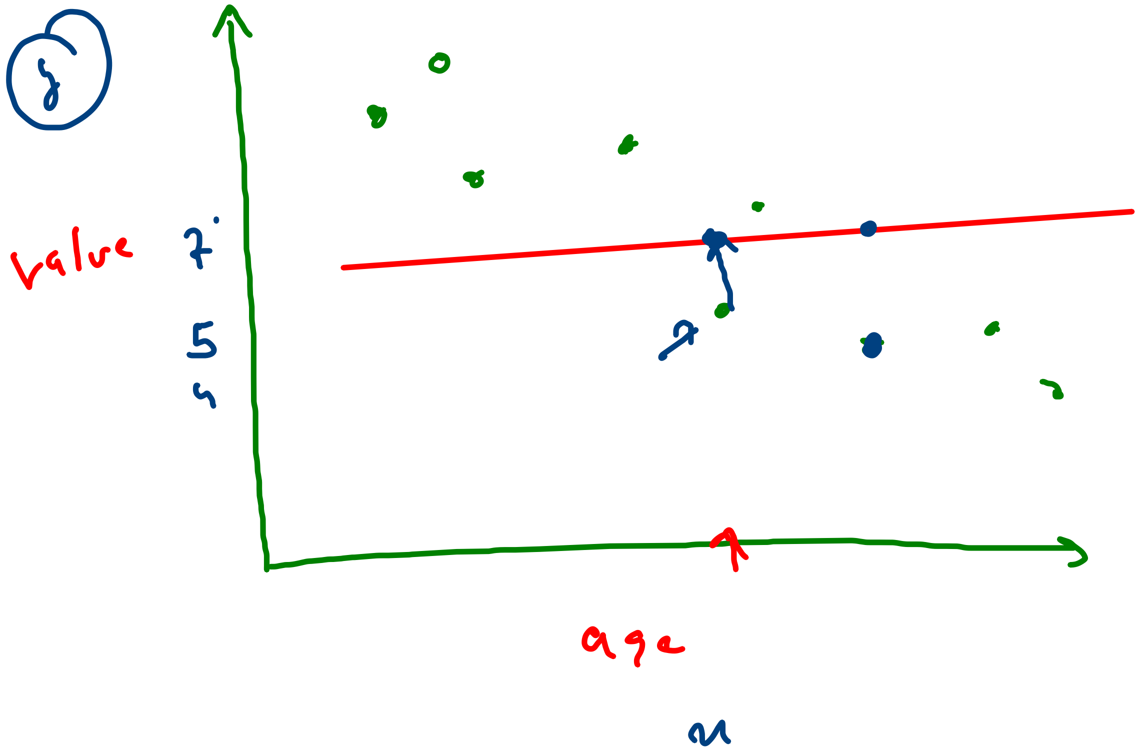


LR → $x \rightarrow y$ Linear relation

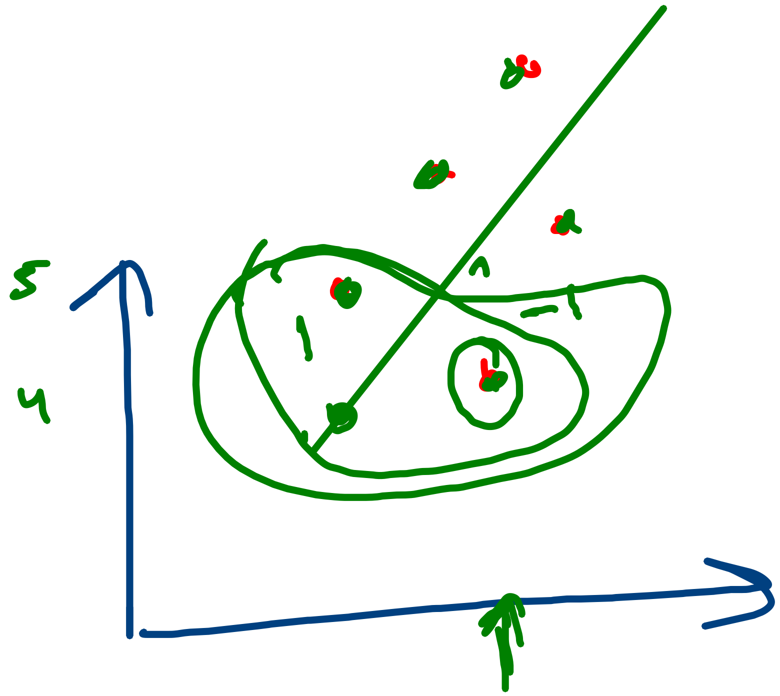
Cost function $= \sum |y_{actual} - y_{predict}|$

$5 - 7 = -2$

$4 - 7.2 = -2.2$



$y = mx + c$



$$\sum |y_{act} - y_{pred}|$$

$$4 - 5 = -1$$

$$\sum 0$$

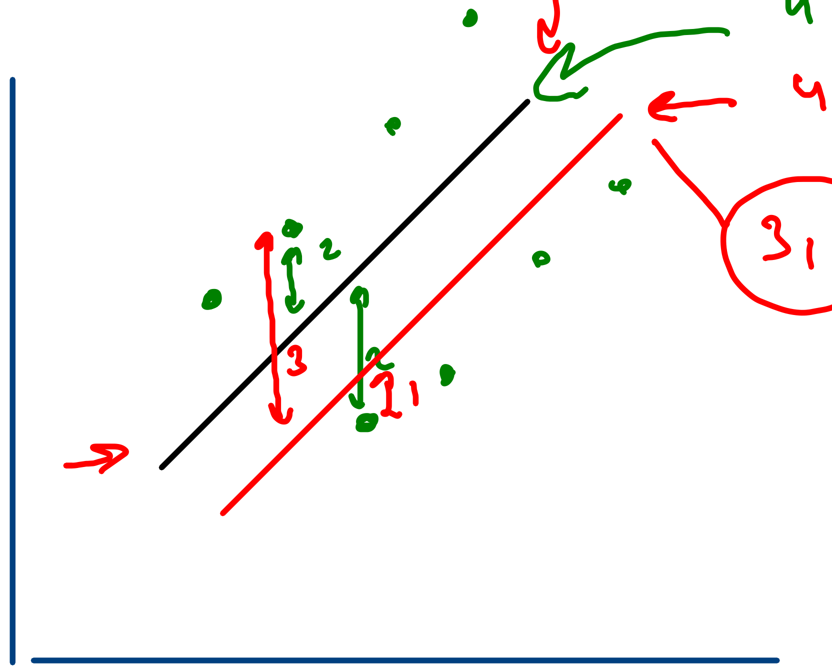
$$5 - 4 = 1$$

Slope, intercept

2,2 → $2^2 + 2^2 = 8$

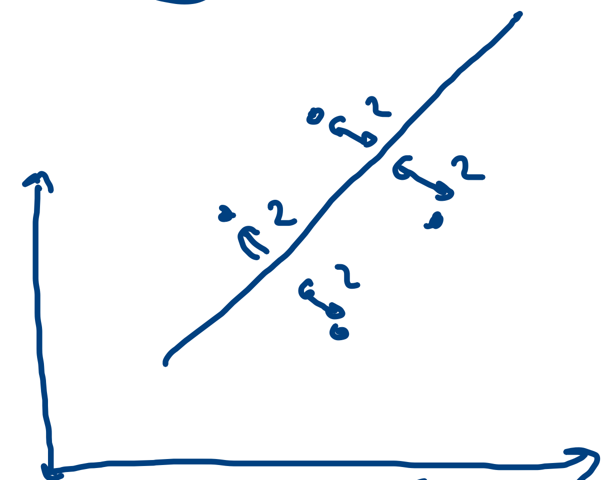
MSE

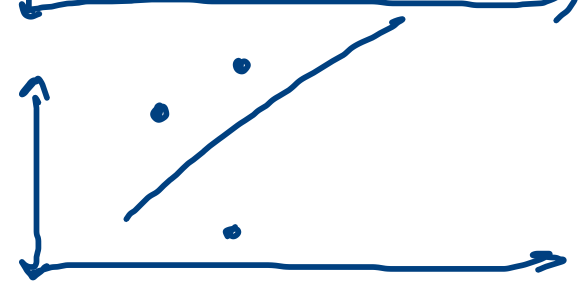
cost function $= \frac{1}{m} \sum \left[y_{actual} - y_{predict} \right]^2$



3,1 → $9 + 1 = 10$

$\frac{8}{4} = 2$





$= \frac{6}{3} = 2$

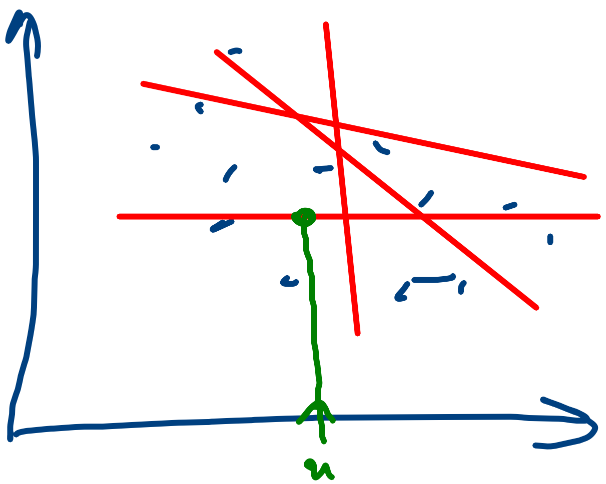
cost function $= \frac{1}{N} \sum \left(y_{actual} - y_{pred} \right)^2$

$f(m, c) = \frac{1}{N} \sum \left(y_a - (mx + c) \right)^2$

line

line $\{ \text{slope}, \text{intercept} \}$

$\{ m, b \}$



best fit $y = mx + c$

* Gradient Descent

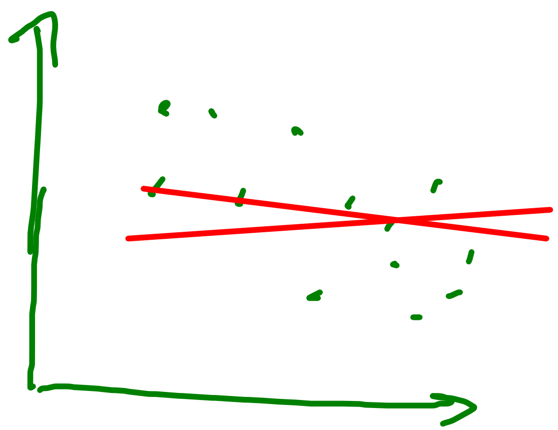
initial line $\{m_1, b_1\}$ 10

$\{m_2, b_2\}$ 8

$\{m_3, b_3\}$ 5

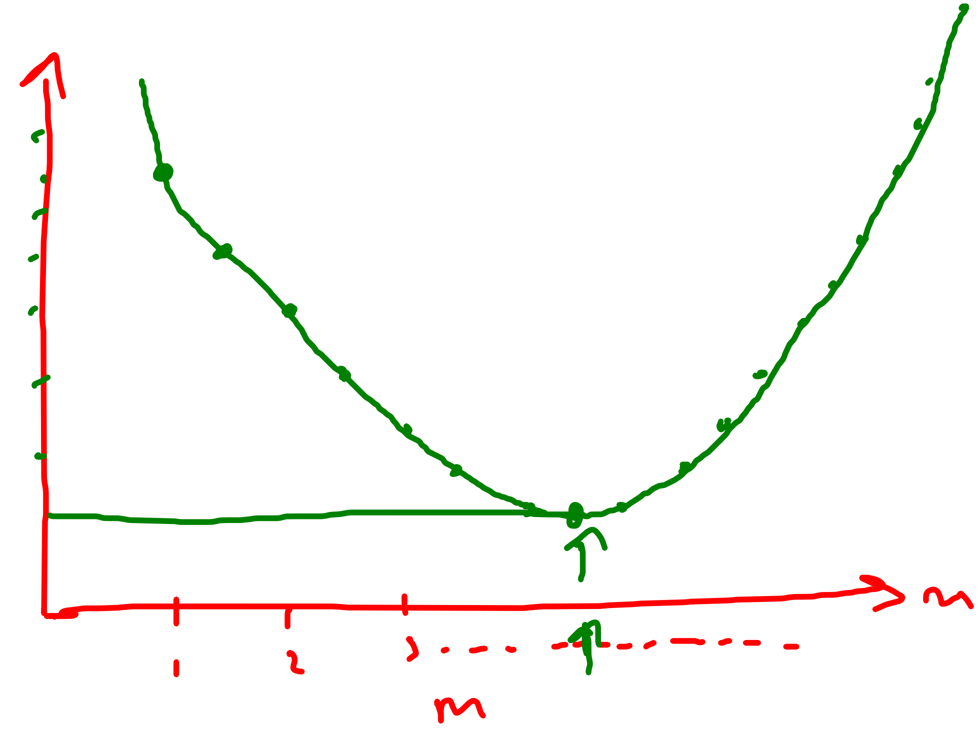
$\vdots \rightarrow$ 2

Gradient Descent



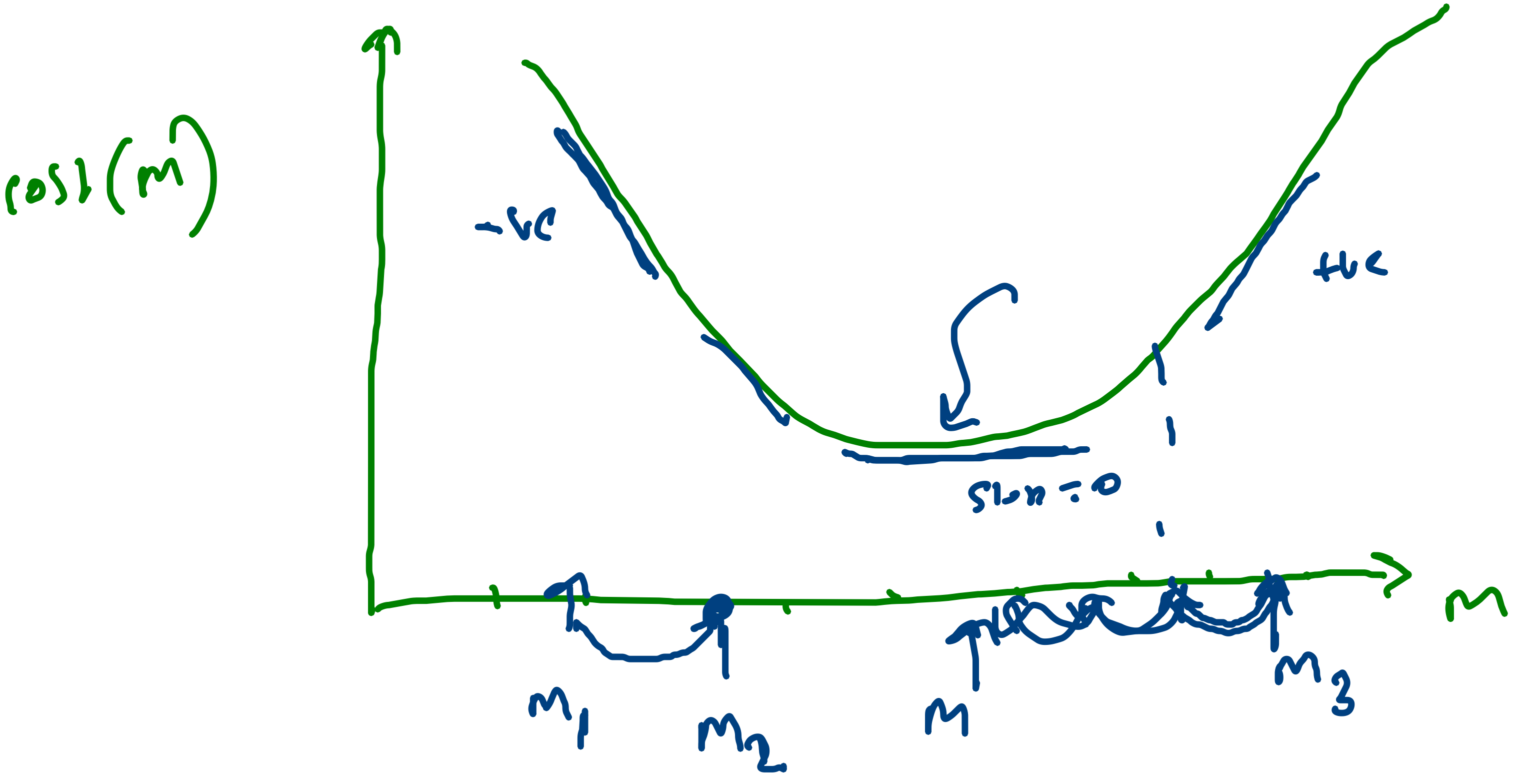
cost (m, c)

$lin = \{m, c\}$



$$\boxed{m_{new} = m_{old} - \lambda \frac{d\,\text{cost}(m)}{dm}}$$

Learning rate



$$m_2 = m_1 - (-ve)$$

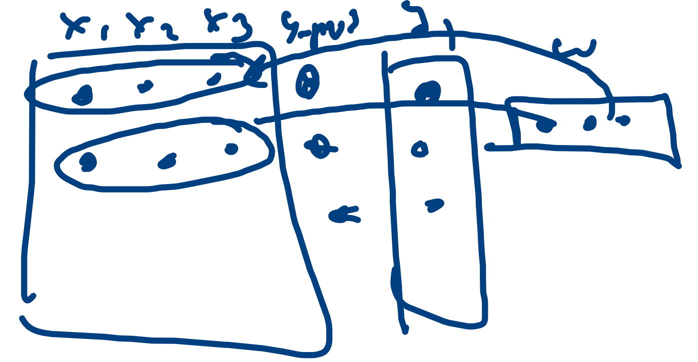
$$m_3 = m_2 - (-ve)$$

$$m_4 = m_3 - (+ve)$$

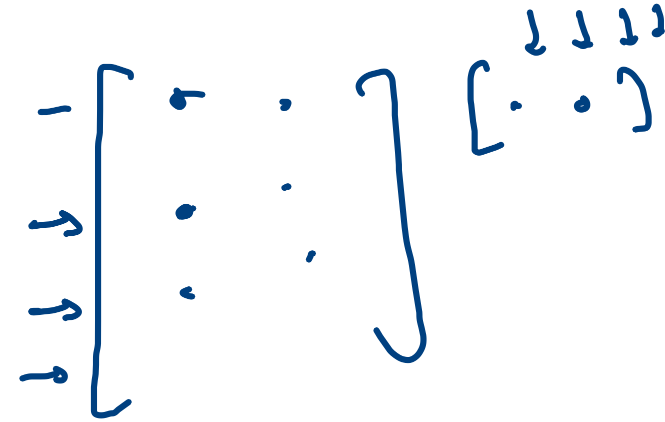
d

$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{dy}{dx} = \frac{d\,\text{cost}(m)}{dm}$$

$$\text{cost} = \frac{1}{N} \sum \left(y_{act} - (mx + b) \right)^2$$



$$\frac{d\,\text{cost}}{dm} = -2 \frac{1}{N} x \sum \left(y_{act} - (mx+b) \right)$$

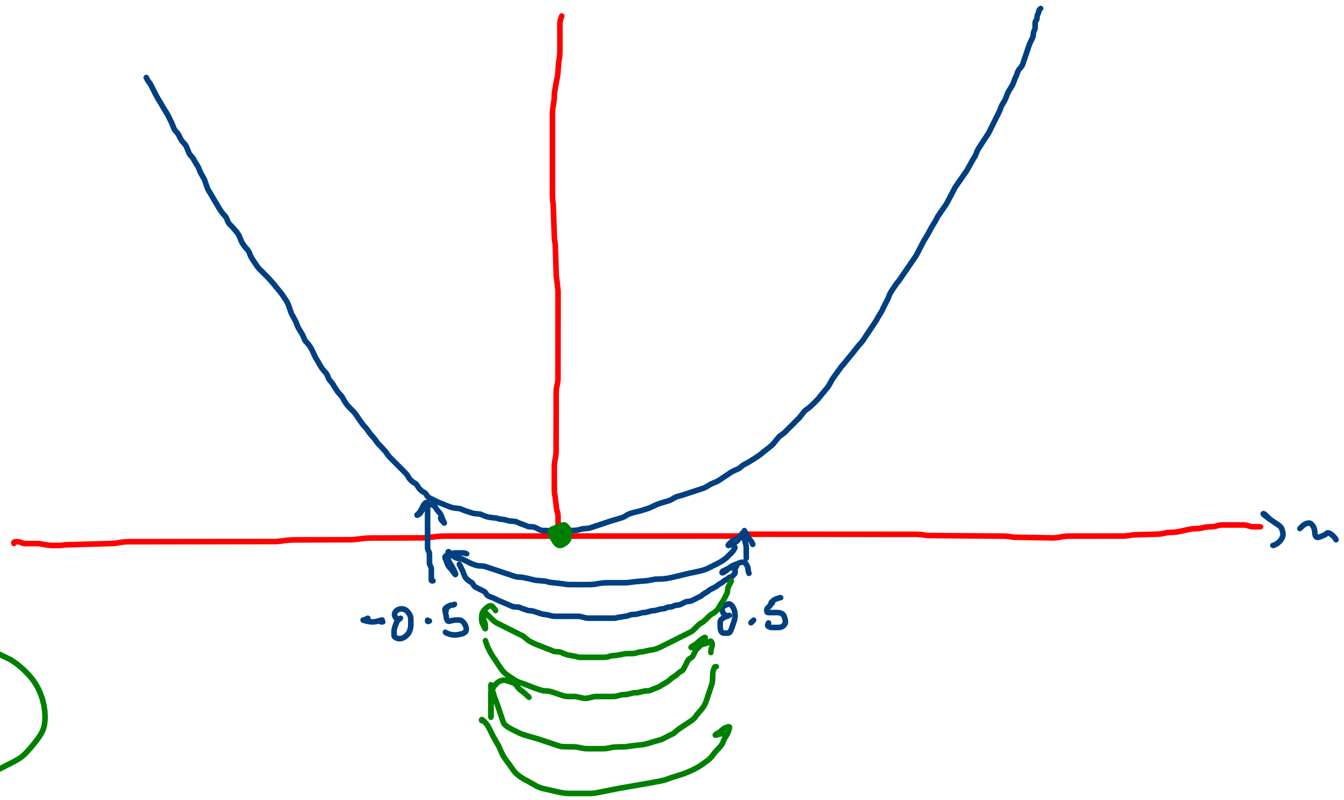


$$\frac{d\,\text{cost}}{db} = -2 \frac{1}{N} \sum \left(y_{act} - (mx+b) \right)$$

$$\boxed{m_{new} = m_{old} - \lambda \frac{d\,\text{cost}(x)}{dm}}$$

$\lambda \times 1$

$\boxed{0.01}$



$y = x^2$

1

$$\frac{dy}{dx} = \frac{d(x^2)}{dx} = \boxed{2x}$$

$x_2 = x_1 - \text{slope}$

$= 0.5 - 1$

$= -0.5$

$x_3 = x_2 - (-1)$

$= -0.5 + 1$

$= 0.5$

1×0.01

0.01

$0.5 - 0.01$

0.49

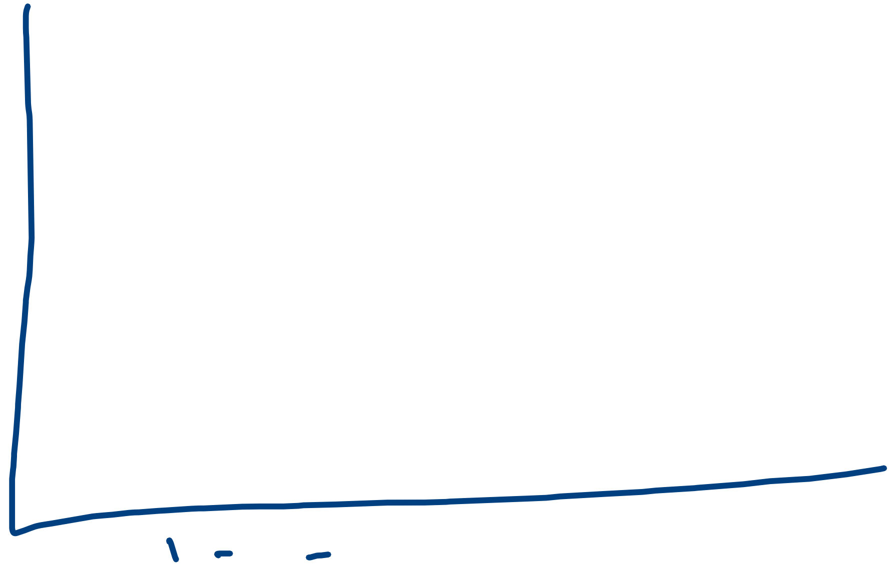
x

[a b c d e]

shape(5)

[[a]
[b]
[c]
[d]]

$\rightarrow (5, 1)$



$\boxed{y = 3x + 5}$

$m = 3$

$b = 5$

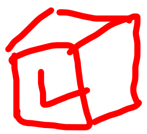
Linear Regression is heavily effected by Outlier's

x_1	x_2	
1	4	5
2	4	9
3	3	5
4	5	6
5	7	8
6	8	9
		6

$x_1 \quad x_2 \quad x_3$

[m_1, m_2, m_3]

$$\boxed{y = m_1x_1 + c}$$



$$y = m_1x_1 + m_2x_2 + c$$

$d \rightarrow$ dimensions

$$y = \boxed{m_1x_1 + m_2x_2 ---- m_dx_d} + c$$

m_0

hyperplane

$$\boxed{y = w^Tx + b}$$

w_0

$$w^T \qquad\qquad x^T$$

$$\begin{bmatrix} a & b & c & d \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} a \times 1 + 2b + 3c + 4d \end{bmatrix}$$

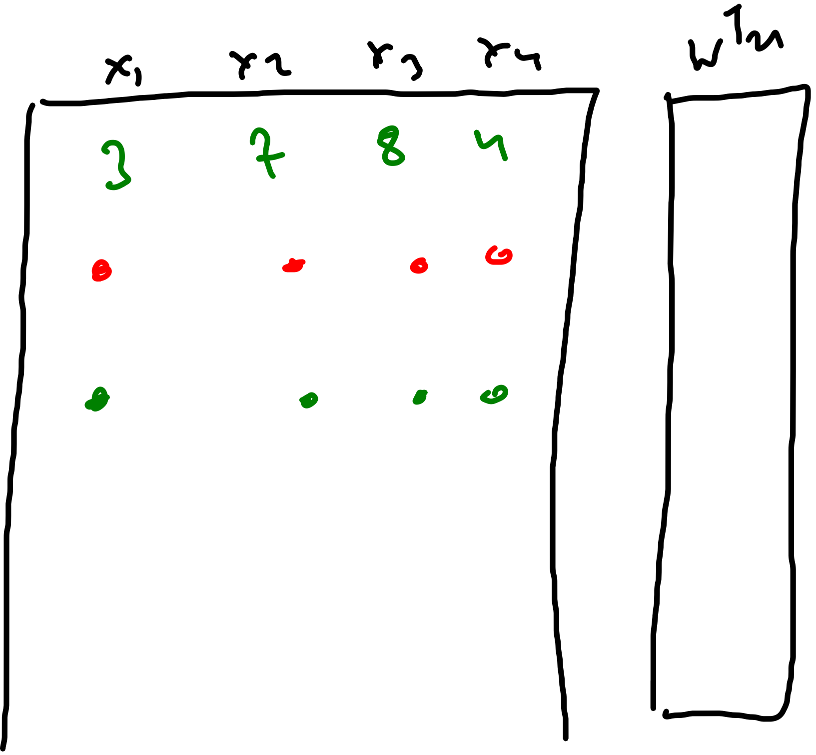
$$\overset{w^T}{\begin{bmatrix} a & b & c & d \end{bmatrix}} \overset{x}{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}}$$

$$w^T x \longrightarrow \begin{bmatrix} 1a + 2b + 3c + 4d \end{bmatrix}$$

vector $\rightarrow \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$

$$x^T \rightarrow \begin{bmatrix} a & b & c & d \end{bmatrix}$$

$$y = w^T x + b$$



$w^T x + b$

y

y_1

y_2

y_3

$$[w_1 \quad w_2 \quad w_3 \quad w_4]$$

$$y_1 = 3w_1 + 7w_2 + 8w_3 + 4w_4 + b$$

$$\begin{bmatrix} a & b \\ a & b \\ a & b \end{bmatrix} \quad [a \quad b]$$

Linear Regression

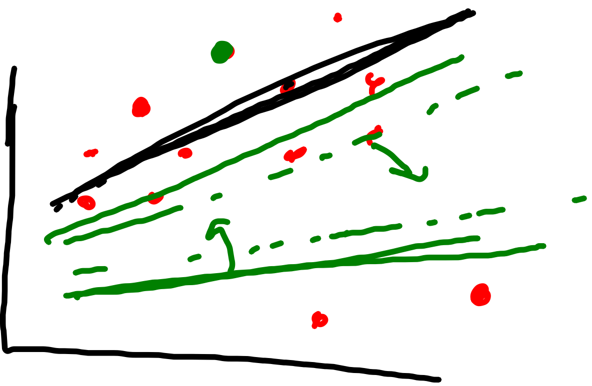
Cost function $= \frac{1}{N} \sum \left(y_{actual} - y_{predicted} \right)^2$ MSE

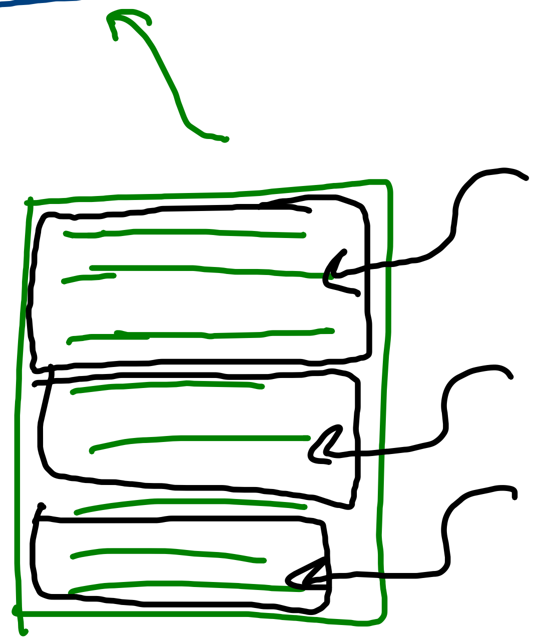
GRADIENT DESCENT $=$ $m_{new} = m_{old} - \lambda \frac{d \,(\text{cost function})}{dm}$

Scratch Implementation

Types of Gradient Descent

X , Y → 10k instance × 1000 ← λ





complete GD

1 • stochastic GD

K • mini batch SGD

$$y = \boxed{m_1 x^2 + m_2 x + c}$$

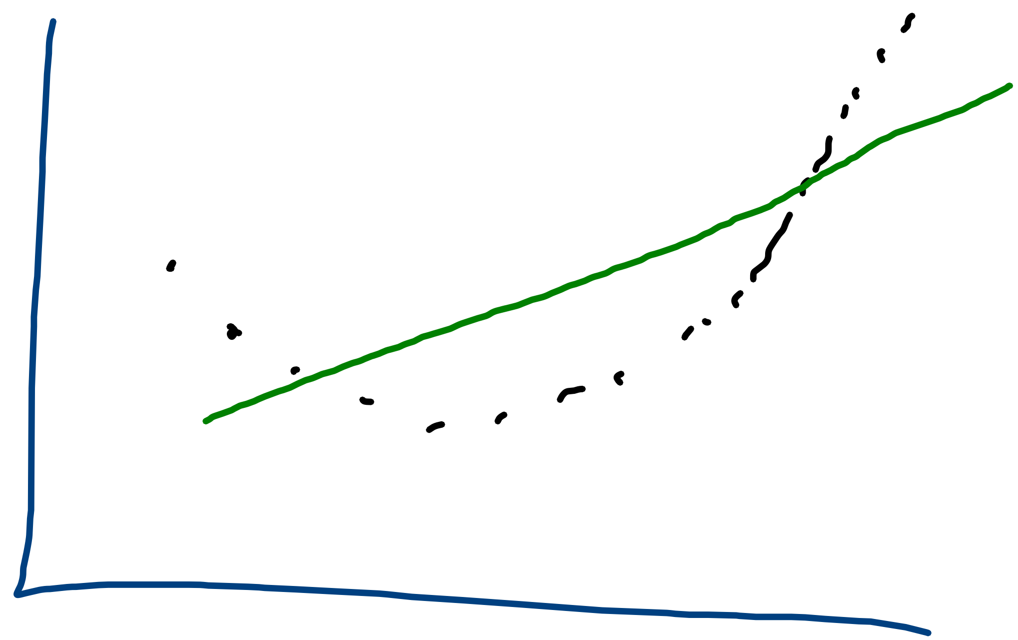
$$y = x^2 + x + \ldots$$

X

x_1 x_1^2

$$\boxed{y = x_1 m_1 + c}$$

$$y = m_1 x_1 + m_2 x^2 + c$$

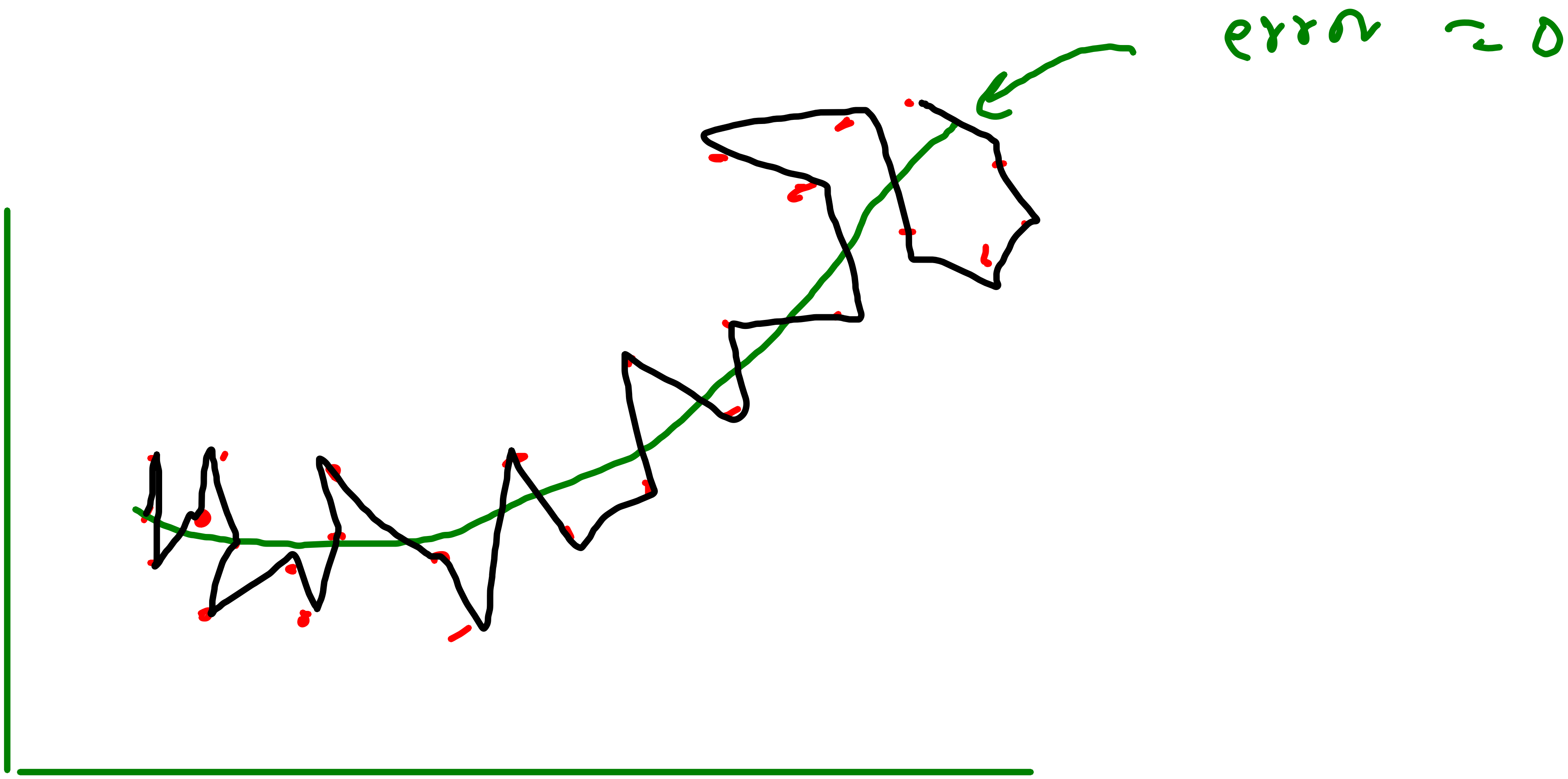


x^1 x^2

x^{10}

degree = 3

d	0	1	2	3
	1	x_1 x_2	x_1^2, x_1^2, $x_1 x_2$	x_1^3, x_2^3, $x_1^2 x_2$, $x_2^2 x_1$
	1			
	1			
	1			
	1			


error = 0

Regularization

L1, L2

Lasso, Ridge

$$\text{Cost function} = \underbrace{\frac{1}{N} \sum \left(y_{actual} - y_{predicted} \right)^2} + \alpha (w)^2$$

$w \to \infty$

$-\infty$

≈ 0

$\alpha \to 0 \to$ overfit

$\to \infty \to$ underfit

no noise

x_1	x_2	x_3	y
•	•	•	8
•	•	•	10
•	•	•	15

$w \to \{2, 3, 4\}$ $\quad y = 2x_1 + 3x_2 + 4x_3 + 10$

noise

x_1	x_2	x_3	y
•	•	•	8.5
•	•	•	10.7
•	•	•	15.6

$w \to \{2.4, 3.2, 4.07\}$

$w \to \{2.2, 3.13, 4.0001\}$

≈ 0

Ridge $\quad \text{Cost function} = \frac{1}{N} \sum \left(y_{actual} - y_{predicted} \right)^2 + \alpha \frac{1}{2} (w)^2 \longrightarrow$ 2 norm

L2

$= 0$

Lasso $\quad \text{Cost function} = \frac{1}{N} \sum \left(y_{actual} - y_{predicted} \right)^2 + \alpha \, |w| \longrightarrow$ 1 norm

L1

ElasticNet $\quad \text{Cost function} = \frac{1}{N} \sum \left(y_{actual} - y_{predicted} \right)^2 + \alpha \, \text{Ridge} + (1-\alpha) \, \text{Lasso}$

0 . . . - - - - 1

Lasso $\qquad\qquad$ Ridge

x	$2x$
2	4
3	6
4	8
7	14

$$y = 8x + 4x^2$$

$$y = 3(2x) + 2x$$

$$4x + 2(2x)$$

height age gender lodhmarly weight