Age	wt	ht	salary
20	85	150	20000
50	74	160	30000
80	68	170	40000
35	54	180	50000

1) Standardization → Z-score standardization

ht BMI wt

~~1187~~
158

20
25
30
40

$$x_{scaled} = \frac{x_i - \bar{x}}{\sigma}$$

$$= \frac{40 - 30}{5}$$

$$= 2$$

$\bar{x} = 30$

$\sigma = 5$

$$\frac{21 - 30}{5}$$

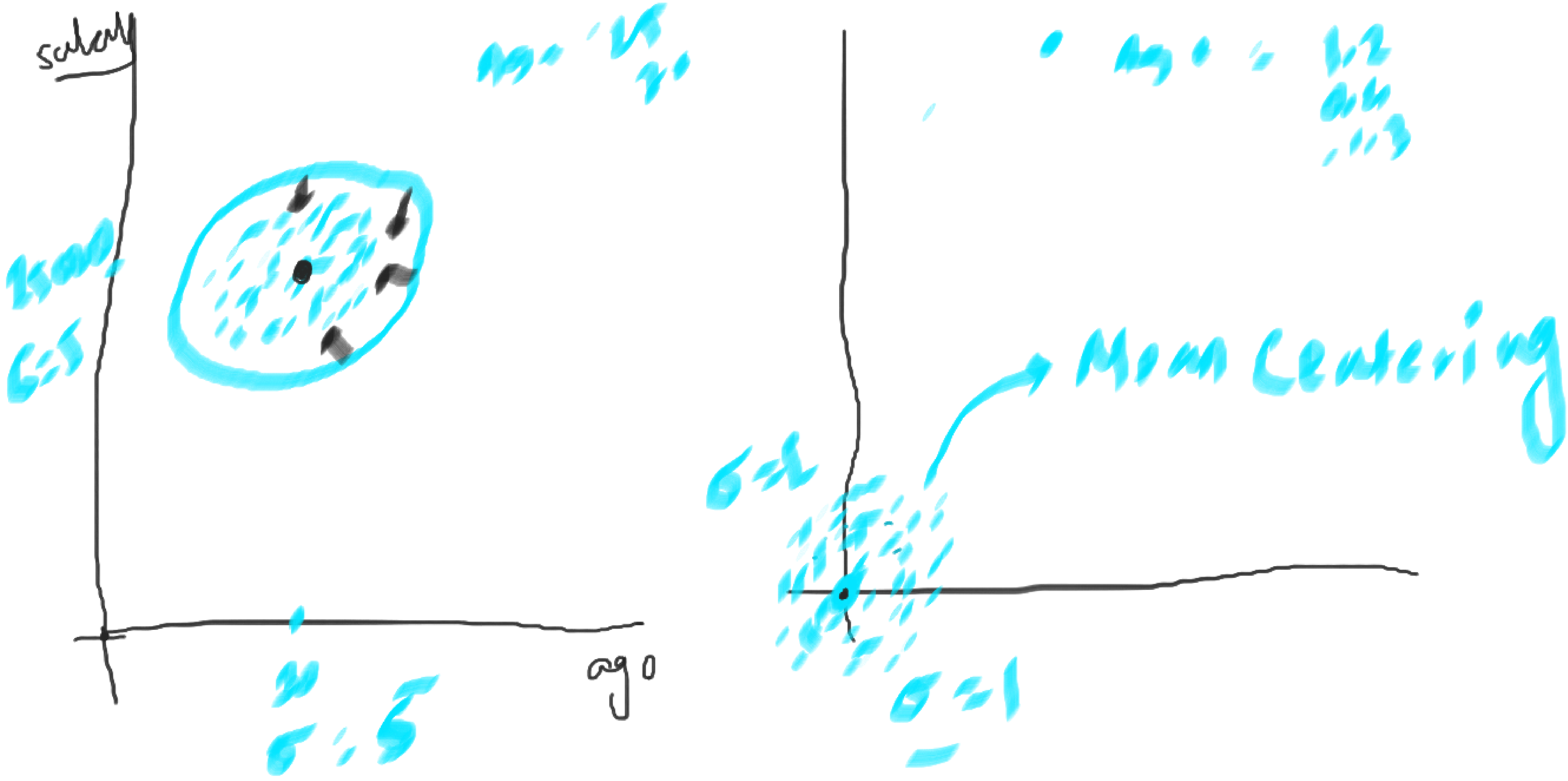
$$= \frac{-3}{5}$$

$$= -0.6$$

$$\frac{25 - 30}{5}$$

$$= -1$$

(−1 to 1)


salary
age
30
σ = 5
σ = 1
σ = 1
Mean Centering

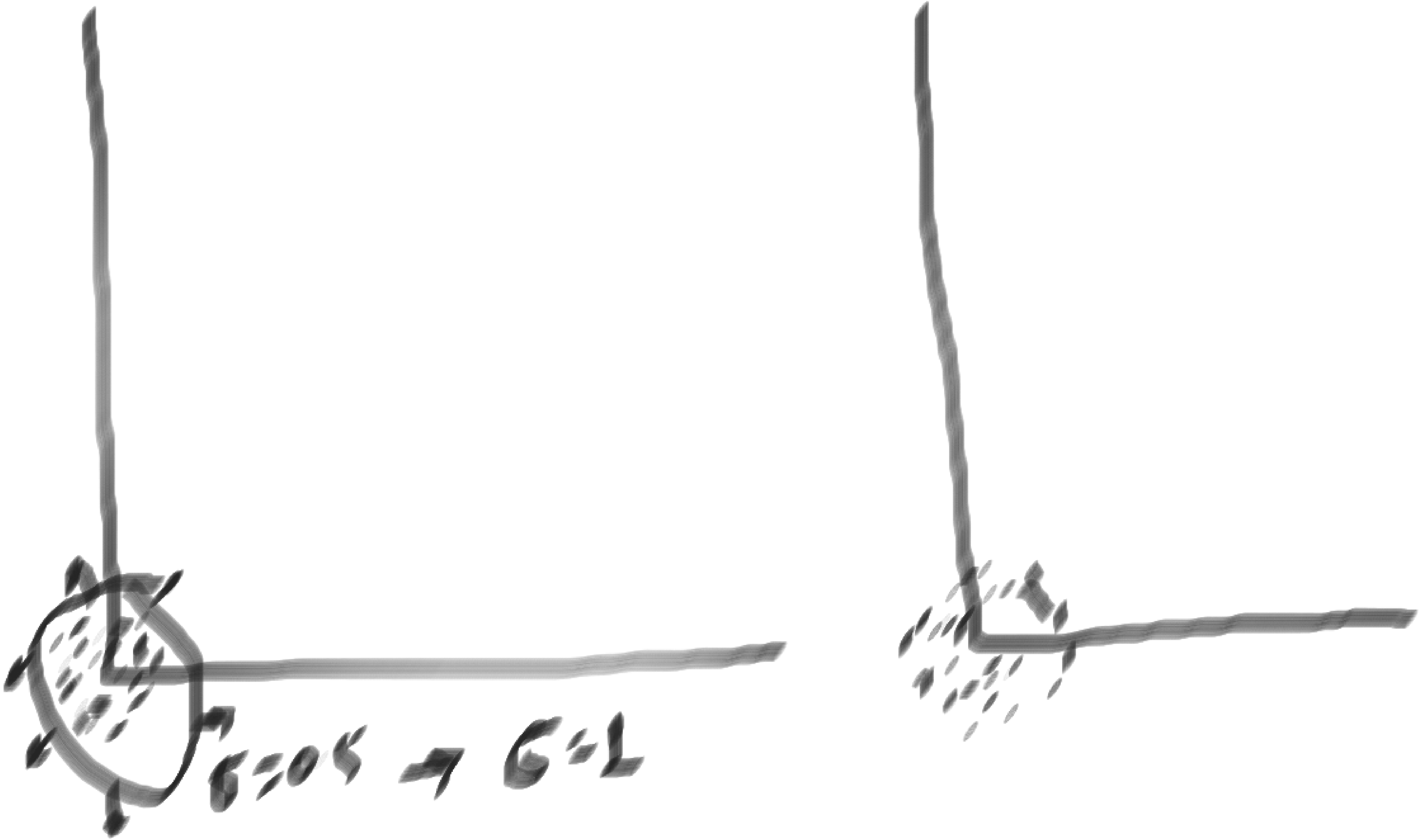
$27-30$

$= \underline{\underline{-ve}}$

$= \underline{\underline{0}}$

$35-30$

$\underline{\underline{+ve}}$


σ=05 → σ=1

When std is more than 1 >> squeez $= \sigma = 5$

std = 1

when std is less than 1 >> expand $= \sigma = 0.5$

std = 1

Age	Salary	Age'Scaled
27	10000	0.7
15	18000	-0.8
		-0.8
12	7000	
13	6000	
18	4000	
29	2000	
32	26000	1.2
27	30000	

$m = 20$ $\sigma = 10$

$m = 0$

$\sigma = 1$

$$x_{std} = \frac{x_i - \bar{x}}{\sigma}$$

$$= \frac{27 - 20}{10}$$

$$= 0.7$$

$$x_{12} = \frac{12 - 20}{10}$$

$$= -0.8$$

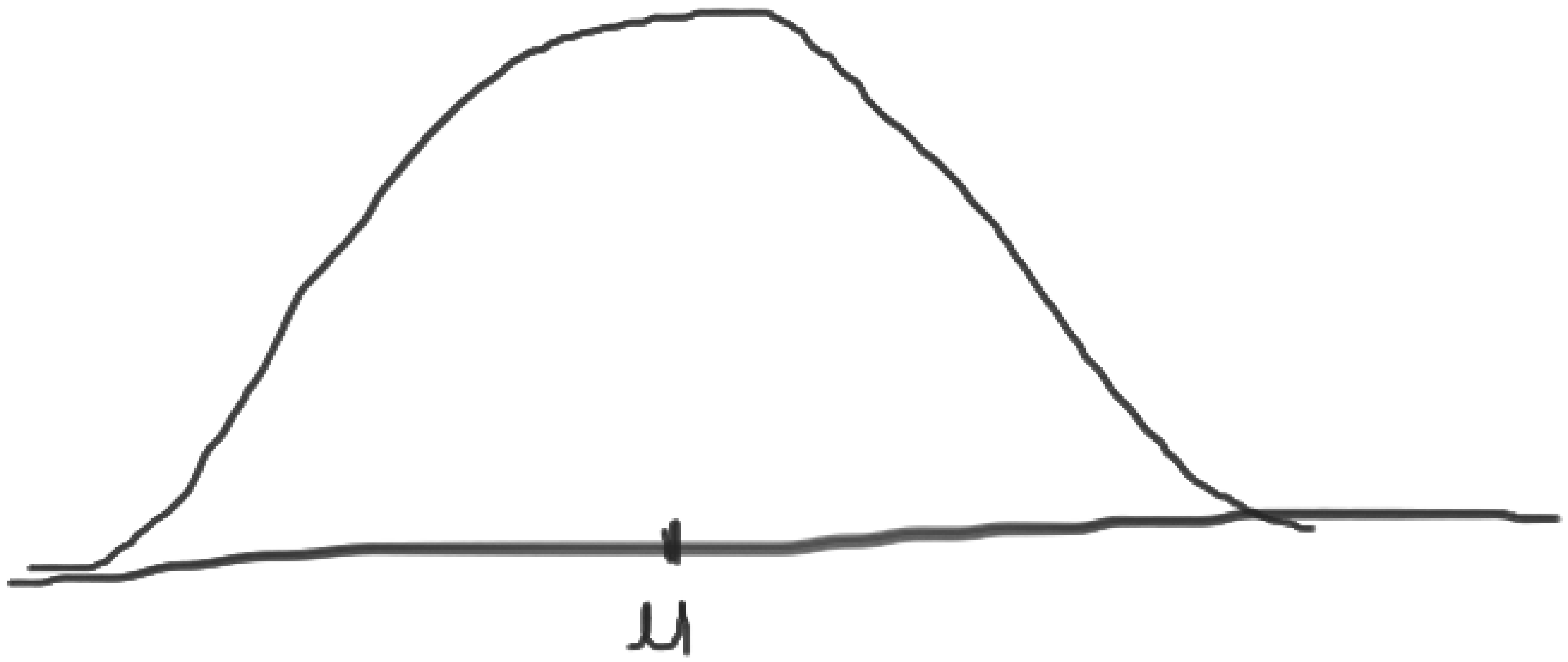
$$\frac{32 - 20}{10} = 1.2$$

Linear = 0.67 → 0.75

ANN = 0.55 → 0.75

Tree = 0.78 ⇒ 0.78

normal = <u>Gaussian</u>




Outlier

Normalization

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

$$x_{14} = \frac{14 - 14}{27 - 14} = \frac{0}{1} = 0$$

$\frac{5}{5}$ $\frac{3}{3}$: 1

$$27 = \frac{27 - 14}{27 - 14} = 1$$

Age	Salary
27	200
15	50
14	40k
18	30k
22	20k
26	21k
21	15k
20	18k
17	24

$x_{min} = 14$ $x_{max} = 27$

$$\frac{22-14}{27-14} = \frac{8}{13} = \underline{\underline{0.61}}$$

$$\frac{14-14}{\,} = \underline{\underline{0}}$$

Distance based algo

euclidean di = $\sqrt{a^2+b^2}$ =

	Age	sal
	25	6000
A	30	7000
B	32	8000
	35	9000
	40	10000

$$\min = \frac{25-25}{40-25} = 0$$

$$\max = \frac{40-25}{40-25} = 1$$

0

0.33

$$\frac{30-25}{40-25} = \frac{5}{15} = 0.33$$

0.13

$$\frac{32-25}{40-25} = \frac{7}{15} = 0.46$$

$$\sqrt{5^2+2^2} = \sqrt{29} = 5.5$$

$$\sqrt{0.33^2+0.13^2} = \sqrt{0.91} = 0.35$$

Max abs

$$x_i = \frac{x_i}{|X_{max}|} = \text{Max Abs Scaler}$$

When we have data with more Zeros

Mean normalisation

$$x_i = \frac{x_i - x_{mean}}{x_{max} - x_{min}}$$

>> mean Centering

standardization

$[-1 \quad 0 \quad 1]$

Robust Scaler

$$x_i = \frac{x_i - x_{median}}{IQR} \quad (Q_3 - Q_1)$$ >> Robust Scaler.

>> Outliers >> effective.

2000000 210000 100000

Age	Prev	prog	Salary	pr	
20	22	2	2000	22000	2000
30	23	1	3000		1000
40	39	1	4000		[illegible]
50	51	1	5000		
60	64	4	6000		4000

16

160000

1,60,00000

lne $= \infty$

2∞

$\frac{0}{1} = 0$

$\frac{1}{0}$ $\frac{2\infty}{0} = \infty$

$\infty - \infty$