Assignment

Dataset

~~Sl.~~	Price
01.	110
02.	105
03.	115
04.	120
05.	110
06.	130
07.	150
08.	100
09.	105

Find:

- Normalization
- Standardization
- Log transformation
- Robust scaler
- Max Absolute scaler

Normalization:

$$\boxed{x_{new} = \frac{x_i - \min(x)}{\max(x) - \min(x)}}$$

$$\frac{110-105}{150-105} = 0.11$$

$$\frac{105-105}{150-105} = 0$$

$$\frac{115-105}{150-105} = 0.22$$

$$\frac{120-105}{150-105} = 0.33$$

$$\frac{110-105}{150-105} = 0.11$$

$$\frac{130-105}{150-105} = 0.55$$

$$\frac{150-105}{150-105} = 1$$

$$\frac{100-105}{150-105} = -0.11$$

$$\frac{105-105}{150-105} = 0$$

	New_Price
01.	0.11
02.	0
03.	0.22
04.	0.33
05.	0.11
06.	0.55
07.	1
08.	-0.11
09.	0

standard scaler:

Standardization:

01. 110
02. 105
03. 115
04. 120
05. 110
06. 130
07. 150
08. 100
09. 105

$\Sigma = 1045/9$

$= 116.11$

$$\boxed{x_{mean}/\mu = 116.11}$$

$$\boxed{x_{new} = \frac{x_i - x_{mean}}{SD}}$$

$$SD, \sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$$

$$= \sqrt{\frac{(110-116.11)^2 + (105-116.11)^2 + (115-116.11)^2 + (120-116.11)^2 + (110-116.11)^2 + (130-116.11)^2 + (150-116.11)^2 + (100-116.11)^2 + (105-116.11)^2}{9}}$$

$$= \sqrt{\frac{1938.89}{9}}$$

$$= 14.678$$

$$\frac{110-116.11}{14.678} = -0.416$$

$$\frac{105-116.11}{14.678} = -0.756$$

$$\frac{115-116.11}{14.678} = -0.075$$

$$\frac{120-116.11}{14.678} = 0.265$$

$$\frac{110-116.11}{14.678} = -0.416$$

$$\frac{130-116.11}{14.678} = 0.946$$

$$\frac{150-116.11}{14.678} = 2.308$$

$$\frac{100-116.11}{14.678} = -1.097$$

$$\frac{105-116.11}{14.678} = -0.756$$

New-Price

01. −0.416
02. −0.756
03. −0.075
04. 0.265
05. −0.416
06. 0.946
07. 2.308
08. −1.097
09. −0.756

Log Transformation:

		New_Price
01.	110 =	2.041
02.	105 =	2.021
03.	115 =	2.060
04.	120 =	2.079
05.	110 =	2.041
06.	130 =	2.113
07.	150 =	2.176
08.	100 =	2
09.	105 =	2.021

$\log_{10}(\text{price}) = ?$

Max Absolute scaler:

$$\boxed{x_{scaled} = \frac{x}{\max(x)}}$$

$\frac{110}{150} = 0.733$

$\frac{105}{150} = 0.7$

$\frac{115}{150} = 0.767$

$\frac{120}{150} = 0.8$

$\frac{110}{150} = 0.733$

$\frac{130}{150} = 0.867$

$\frac{150}{150} = 1$

$\frac{100}{150} = 0.667$

$\frac{105}{150} = 0.7$

	New_Price
01.	0.733
02.	0.7
03.	0.767
04.	0.8
05.	0.733
06.	0.867
07	1
08.	0.667
09.	0.7

In Generally, we use
- Normalization
- Standardization

- ~~Alo~~ Standard scaler gives good performance most of the cases, then others.

- If outlier exist in data, then use log transformation / Robust scaler.

Robust scaler:

$$x_{scale} = \frac{x_i - x_{med}}{x_{75} - x_{25}}$$

$x_{25} = -0.333$
$x_{50} = 0.0$
$x_{75} = 0.667$

Robust scaler

01. 110
02. 105
03. 115
04. 120
05. 110
06. 130
07. 150
08. 100
09. 105

IQR = 75th Quantile − 25th Quantile

$$x_{scale} = \frac{x_i - x_{med}}{IQR}$$

IQR = 0.667 − (−0.333)
IQR = 1

$\boxed{x_{med} = 110}$

$$\frac{110-110}{1} = 0$$

$$\frac{105-110}{1} = -5$$

$$\frac{115-110}{1} = 5$$

$$\frac{120-110}{1} = 10$$

$$\frac{110-110}{1} = 0$$

$$\frac{130-110}{1} = 20$$

$$\frac{150-110}{1} = 40$$

$$\frac{100-110}{1} = -10$$

$$\frac{105-110}{1} = -5$$

New_price

01.	0
02.	−5
03.	5
04.	10
05.	0
06.	20
07.	40
08.	−10
09.	−5