Assignment 2

(III)

What does normalization and standardization do to the data and the noise? Your answers to all three questions should not exceed 2 pages in total but provide technical descriptions including the use of mathematical notation.

Normalization — set the value from a min, max : 0 to 1

Standardization — set the value σ from the mean
(usually you are -3 to 3 sd from mean)

let σ standard deviation of data.

Normalization steps :

$$X_{normal} = \frac{X_i - min(x)}{max(x) - min(x)}$$

let X be a feature
where $X = \{1, ..., n\}$

This just normalizes the data to be in the set of 0 to 1

ie let $X = [1, 3, 5, 7]$ apply formula and

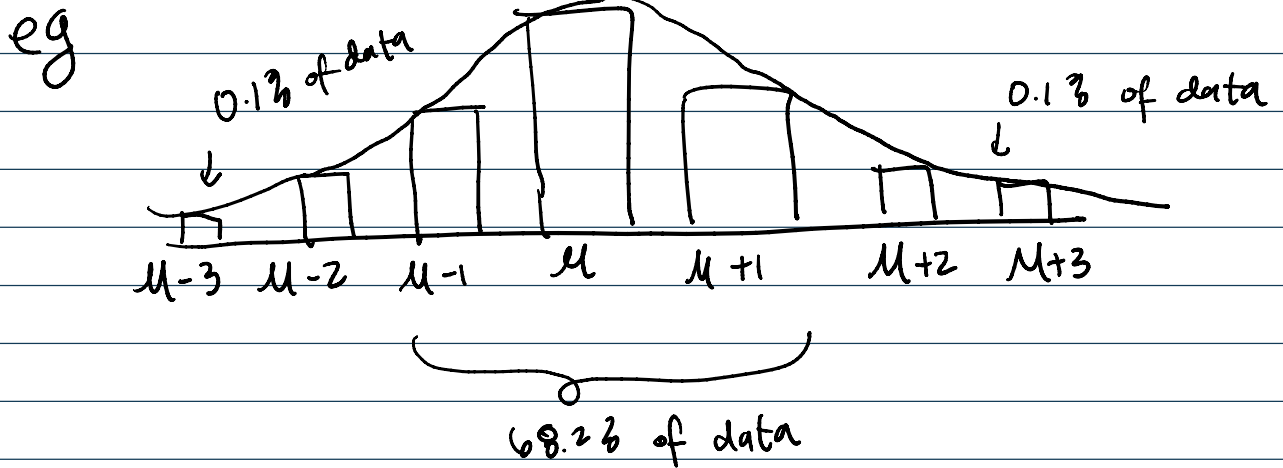$$X_{normal} = [0, 1/3, 2/3, 1]$$ ← notice here that
maxima = 7 is 1
minima = 1 is 0

So range of $X_{normal} = [0, 1]$ exclusively.

Standardization steps

$$X_{standard} = \frac{X - \mu}{\sigma}$$

let μ is mean of dataset
let σ is stdeviation of data

This sets our data into a universe where you must be a certain standard deviation from the mean.

eg

0.1% of data

0.1% of data

$\mu-3$ $\mu-2$ $\mu-1$ μ $\mu+1$ $\mu+2$ $\mu+3$

68.2% of data

The dataset will fall in range of $[-3, 3]$ this value means how far away you are from the mean.

Note: data is not exclusively to $[-3, 3]$ but any thing farther away is very slim that it is improbable.