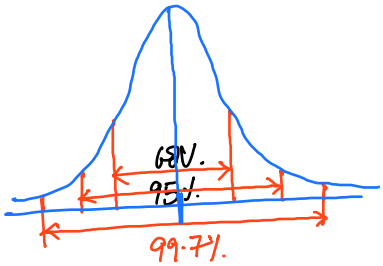
Assignment No 3

Q1 What is a Gaussian distribution empirical formula?

Ans Empirical formula → 68-95-99.9%



Q2 What is the zscore and why it is important?

Z-Score is given by →

$$Z \Rightarrow \frac{x-\mu}{\sigma}$$

* It is important because in ML we have different features, so to bring them on a same scale we used z-score.
* Z-Score is used to convert Normal distribution to standard distribution.

Q3 What is an outlier exactly?

Ans Outliers are the data points which are out in the Range of lower fence and Higher fence

lower Fence → $Q_1 - 1.5\,IQR$

Higher Fence → $Q_3 + 1.5\,IQR$

$IQR \Rightarrow Q_3 - Q_1$

Q4 What are the options for dealing with outliers?

Ans 1) Five Number Summary
↳ Box Plot
↓
Minimum
Q_1 (25 Percentile)
Median
Q_3 (75 Percentile)
Maximum.

Q5

Ans Population (N)

$$\sigma^2 \Rightarrow \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{N}$$

Sample

$$s^2 \Rightarrow \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$$

$(n-1)$ → Bessel correction / Degree of freedom
↳ By using $(n-1)$, it helps not to underestimate the correct value of mean(μ) & Variance (σ)

Q5 Write sample and population variance equation and explain Bessel correction?