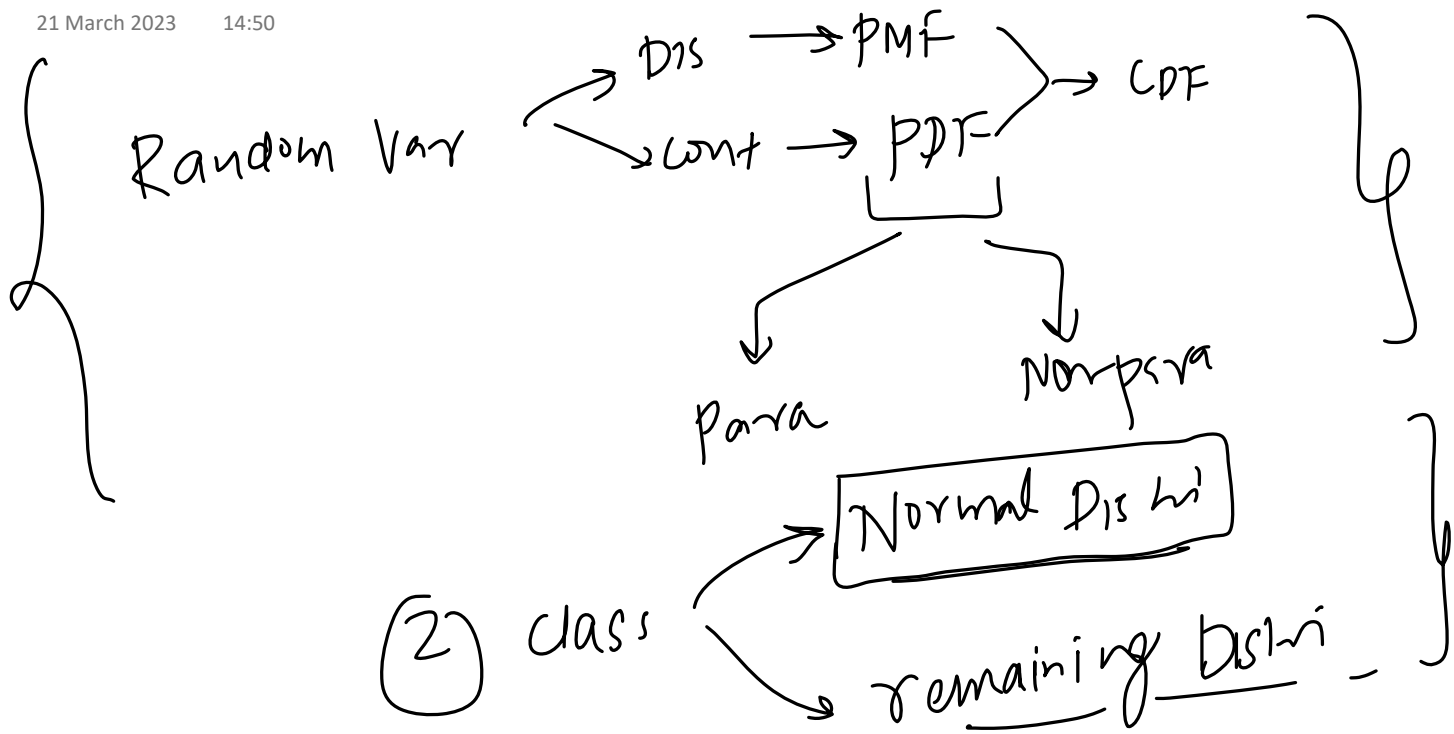


Recap

21 March 2023 14:50



How to use PDF in Data Science

20 March 2023 18:11

PDF →

CDf

PDF

$$P(X=x)$$

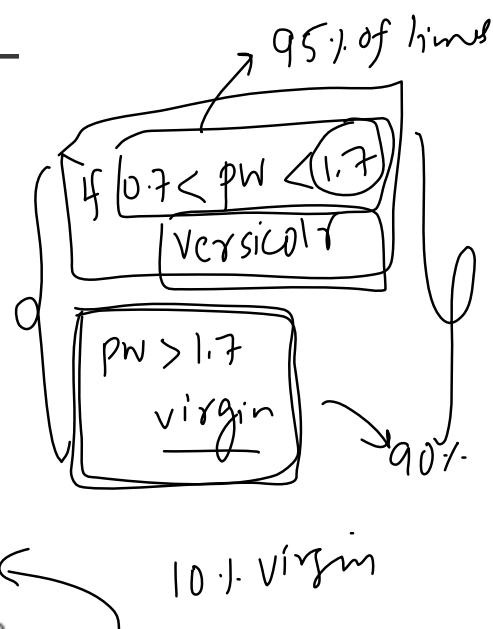
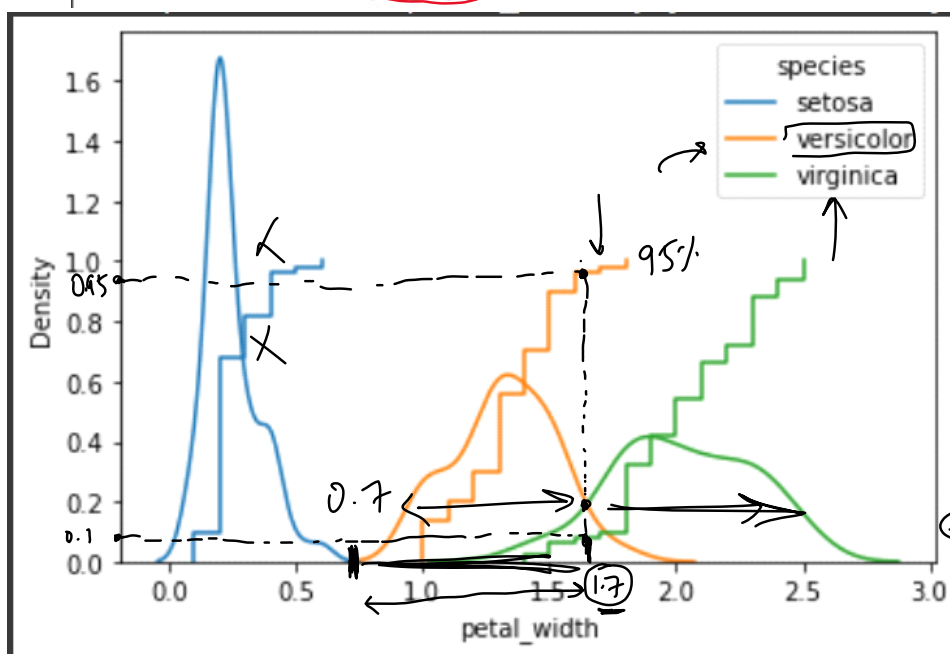
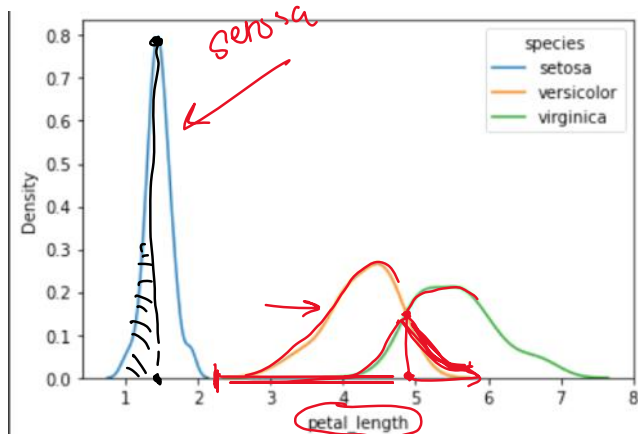
$$2.3 < pl < 5 \rightarrow \text{versicolor}$$

$$P(X \leq x)$$

$$pl < 2.3$$

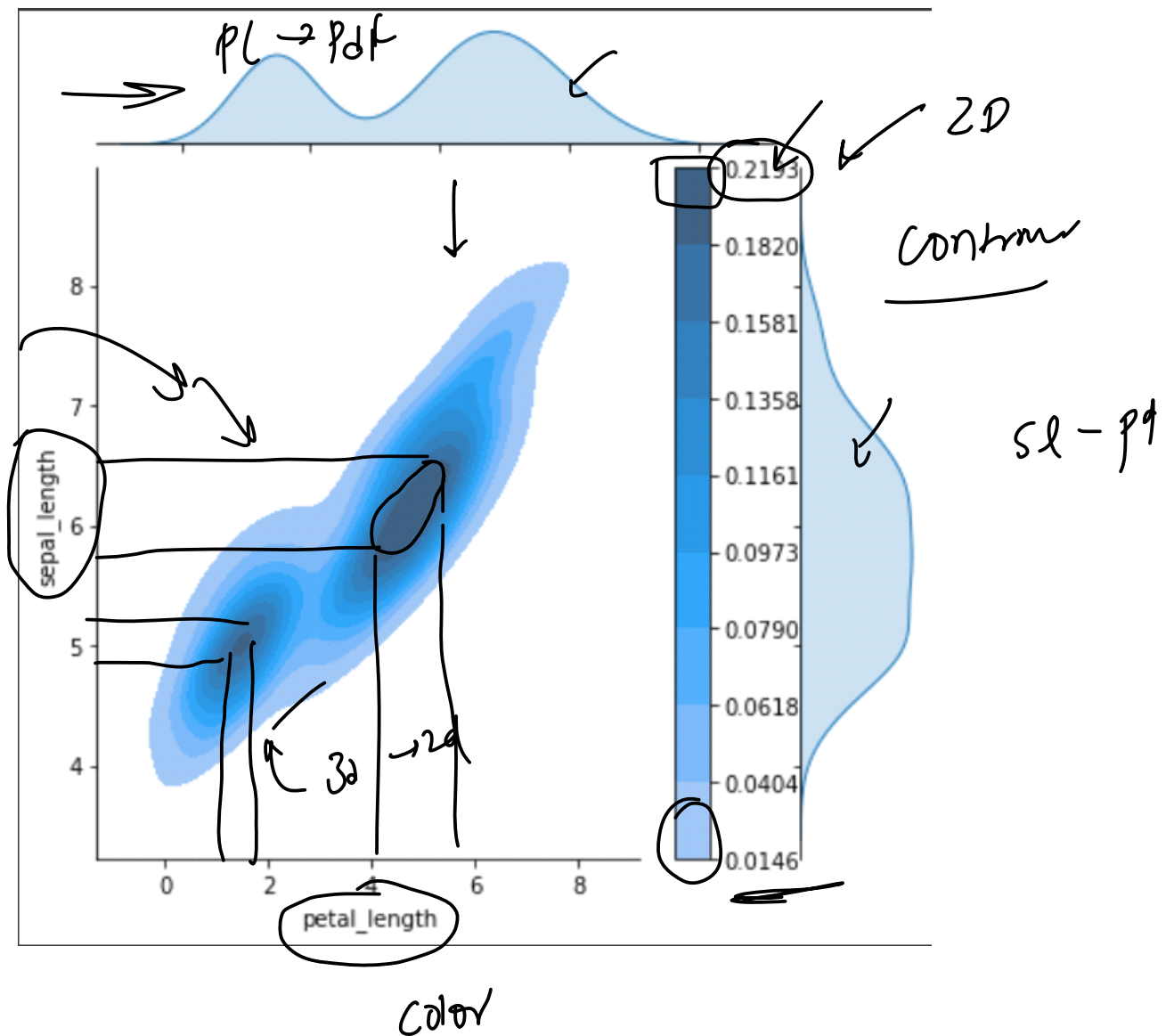
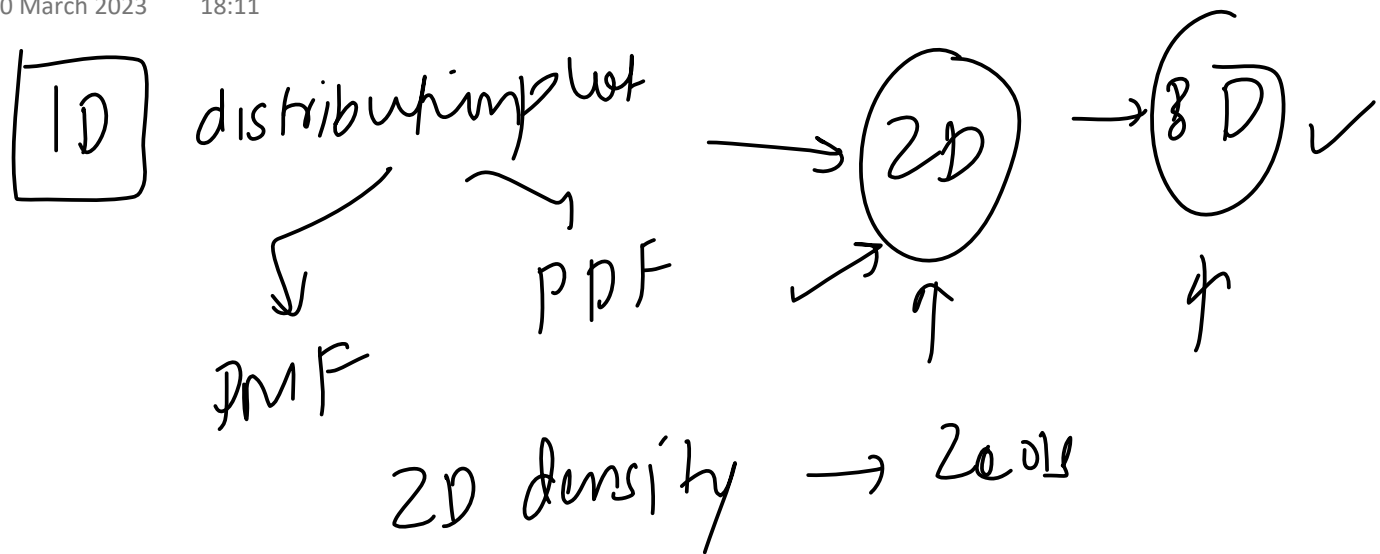
$$> 2.3$$

$$pl > 5 \rightarrow \text{virginica}$$



2D Density Plots

20 March 2023 18:11



20 March 2023 18:06

bell curve

→ pdf



- 4

f

as: $X \sim N(\mu, \sigma)$ $\mu \rightarrow \text{mean}$
 $\sigma \rightarrow \text{std}$

8

No small



data \rightarrow Normal $\rightarrow \boxed{\mu, \sigma}$

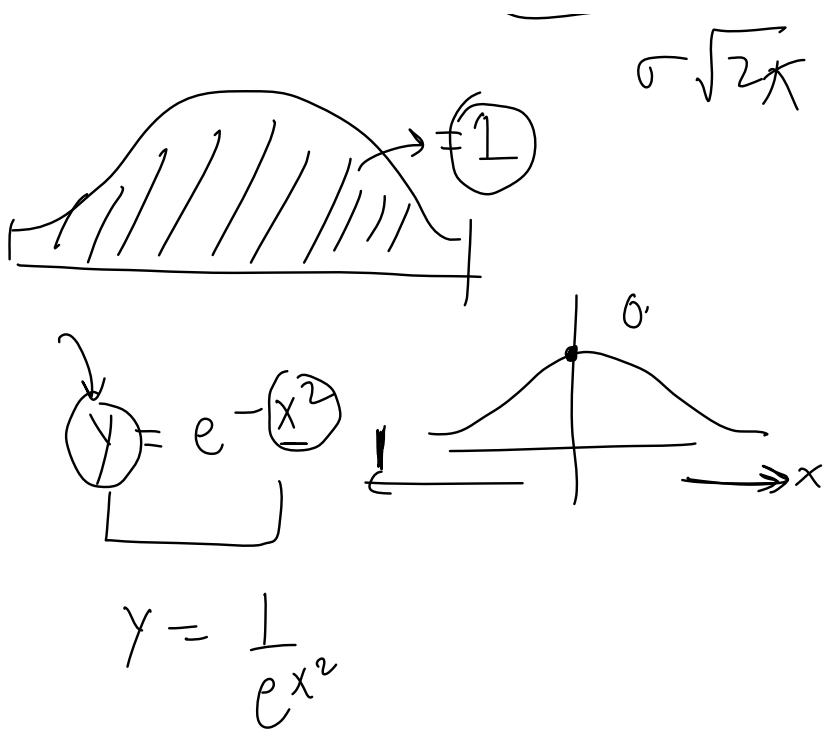
inter \rightarrow $\sigma \sqrt{}$

1

Equation in detail:

Equation in detail:

$$Y = e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$



Standard Normal Variate (Z) → Standard Normal distribution

20 March 2023 18:08

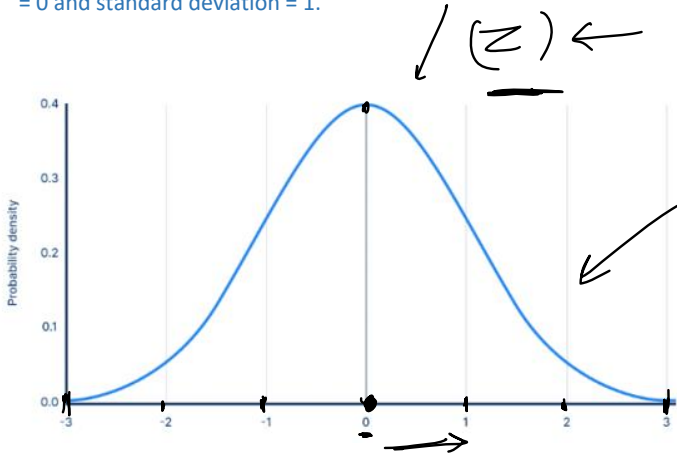
$$X \sim N(\mu, \sigma) \quad \mu=0 \quad \sigma=1$$

$$\downarrow$$

$$Z \sim N(0, 1)$$

What is Standard Normal Variate

A Standard Normal Variate (Z) is a standardized form of the normal distribution with mean = 0 and standard deviation = 1.



Standardizing a normal distribution allows us to compare different distributions with each other, and to calculate probabilities using standardized tables or software.

Equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$X \sim N(5, 2.5)$

How to transform a normal distribution to Standard Normal Variate

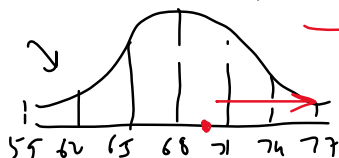
$$X \sim N(29, 14)$$

Refer Python code

Kya Fayda Standardize karne ka?

Suppose the heights of adult males in a certain population follow a normal distribution with a mean of 68 inches and a standard deviation of 3 inches. What is the probability that a randomly selected adult male from this population is taller than 72 inches?

$$X \sim N(68, 3)$$



What are Z-tables

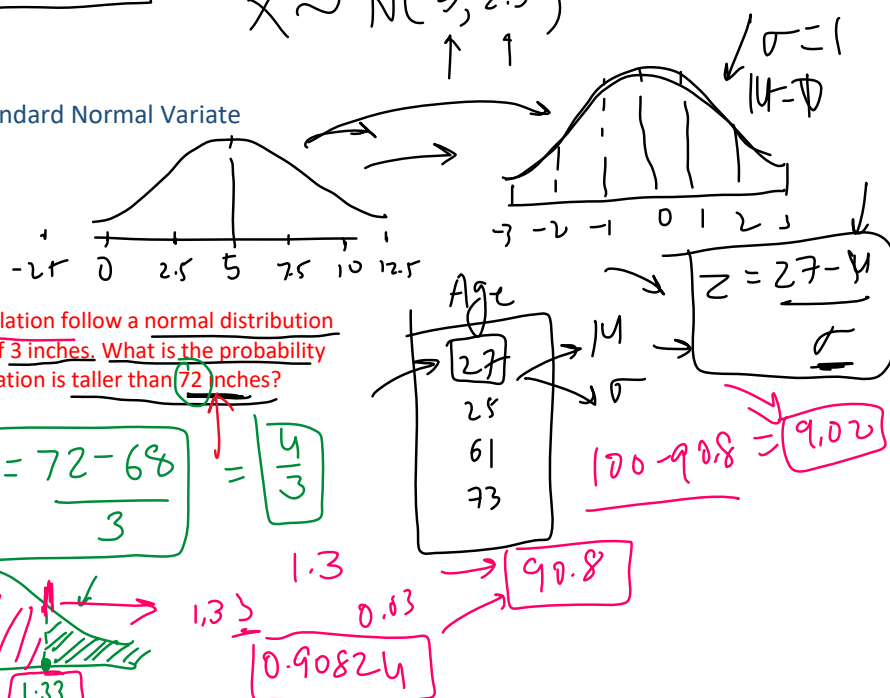
A z-table tells you the area underneath a normal distribution curve, to the left of the z-score

<https://www.ztable.net/>

For a Normal Distribution $X \sim (\mu, \text{std})$ what percent of population lie between mean and 1 standard deviation, 2 std and 3 std?

$$X \sim N(\mu, \sigma) \quad \mu \rightarrow \quad \sigma \rightarrow$$

$$Z = \frac{X - \mu}{\sigma}$$



$$X \sim N(\mu, \sigma)$$

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{\mu - \mu}{\sigma} = 0$$

$$Z = \frac{X - \mu}{\sigma} = \frac{\mu + \sigma - \mu}{\sigma}$$

$$0.9772$$

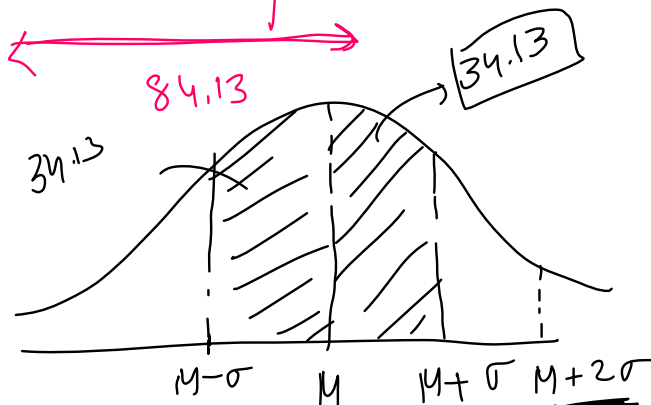
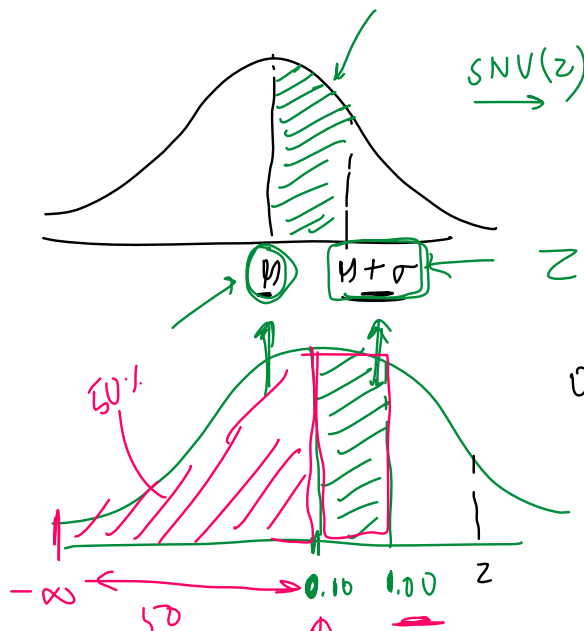
$$97.72\%$$

$$= \frac{\sigma}{\sigma} = 1$$

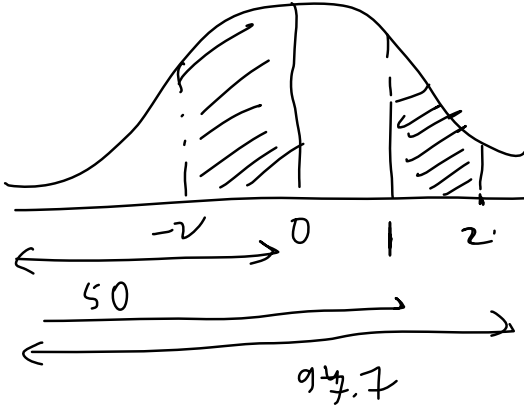
$$84.13 - 50 = 34.13$$

$$-1\sigma \quad +1\sigma$$

$$64.28\%$$



$$Z = \frac{\mu + 2\sigma - \mu}{\sigma} = 2$$

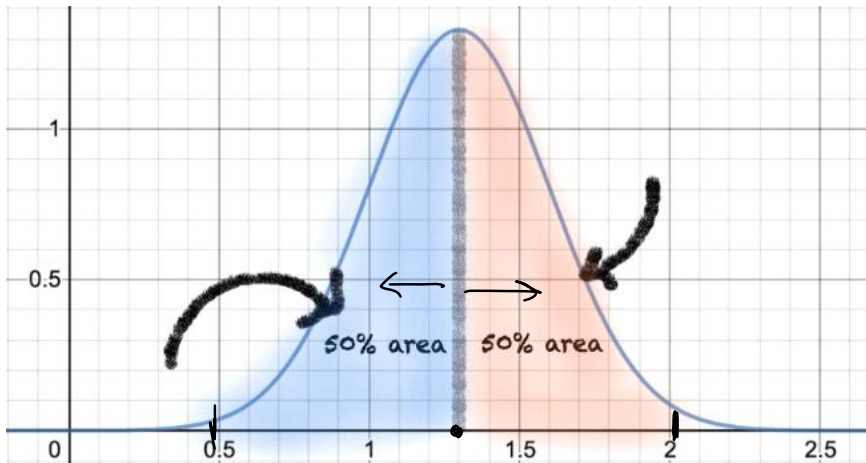


Properties of Normal Distribution

20 March 2023 18:06

1. Symmetry

The normal distribution is symmetric about its mean which means that the probability of observing a value above the mean is the same as the probability of observing a value below the mean. The bell-shaped curve of the normal distribution reflects this symmetry.



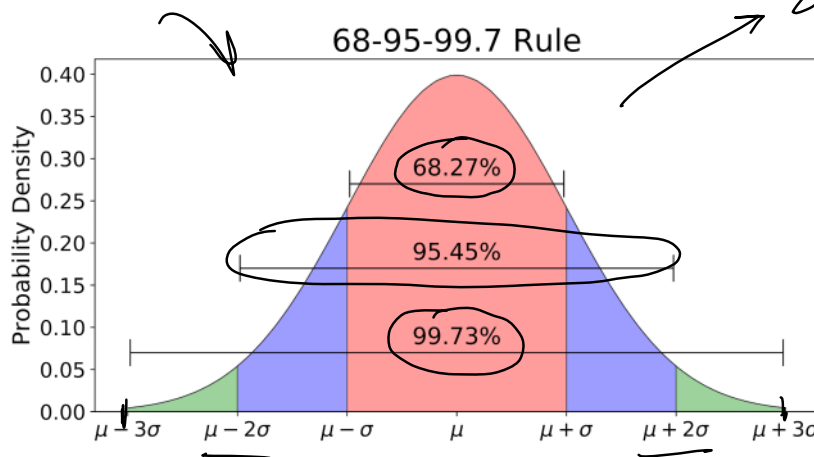
2. Measures of Central Tendencies are equal \rightarrow mean \rightarrow median \rightarrow mode

3. Empirical Rule

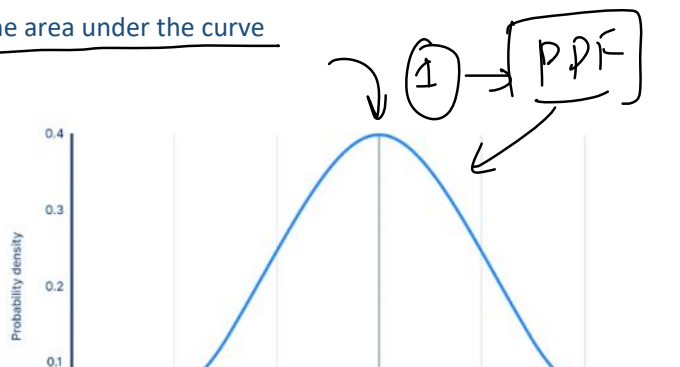
The normal distribution has a well-known empirical rule, also called the 68-95-99.7 rule, which states that approximately 68% of the data falls within one standard deviation of the mean, about 95% of the data falls within two standard deviations of the mean, and about 99.7% of the data falls within three standard deviations of the mean.

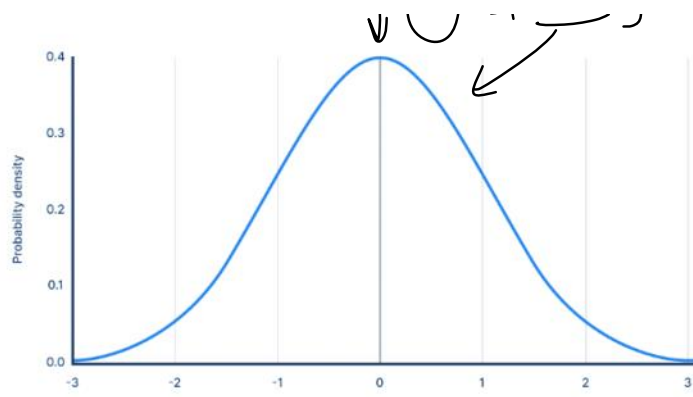
Standard deviation

Z-table



4. The area under the curve



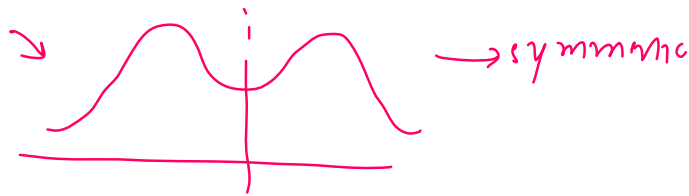


Skewness

20 March 2023 18:07

What is skewness?

A normal distribution is a bell-shaped, symmetrical distribution with a specific mathematical formula that describes how the data is spread out. Skewness indicates that the data is not symmetrical, which means it is not normally distributed.

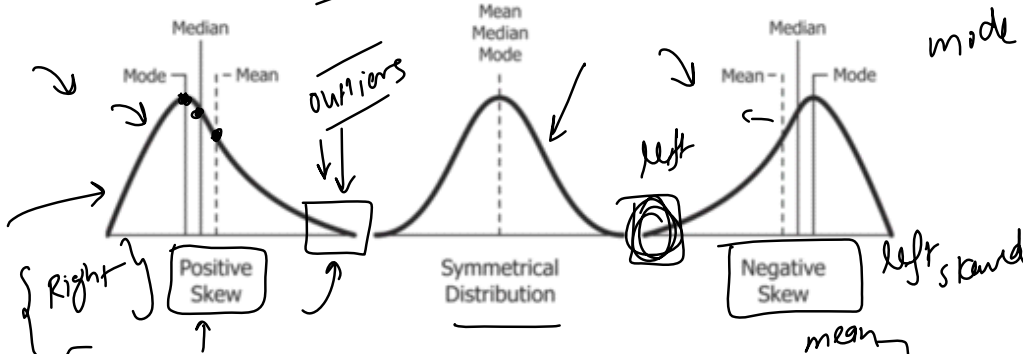


Skewness is a measure of the asymmetry of a probability distribution. It is a statistical measure that describes the degree to which a dataset deviates from the normal distribution.

In a symmetrical distribution, the mean, median, and mode are all equal. In contrast, in a skewed distribution, the mean, median, and mode are not equal, and the distribution tends to have a longer tail on one side than the other.

Skewness can be positive, negative, or zero. A positive skewness means that the tail of the distribution is longer on the right side, while a negative skewness means that the tail is longer on the left side. A zero skewness indicates a perfectly symmetrical distribution.

mode < median < mean
tail event



moment
mode > median > mean
2 moment - variance
3 moment - skewness
4th - kurtosis

The greater the skew the greater the distance between mode, median and mean

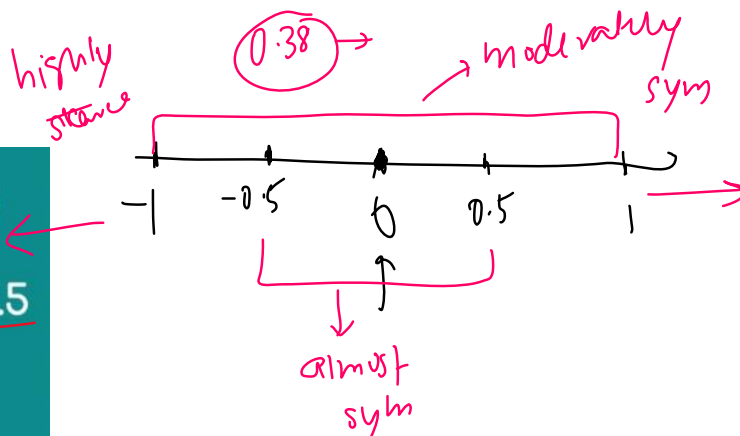
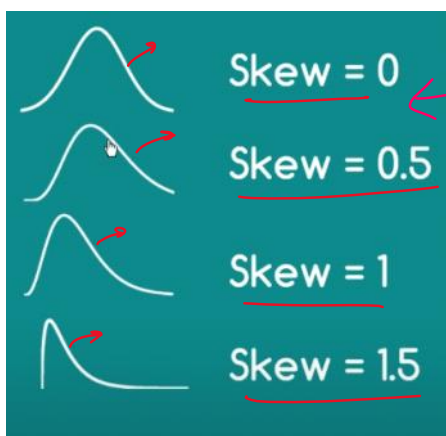
How skewness is calculated?

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{(x - \bar{x})}{s} \right)^3$$

moment
sample skew

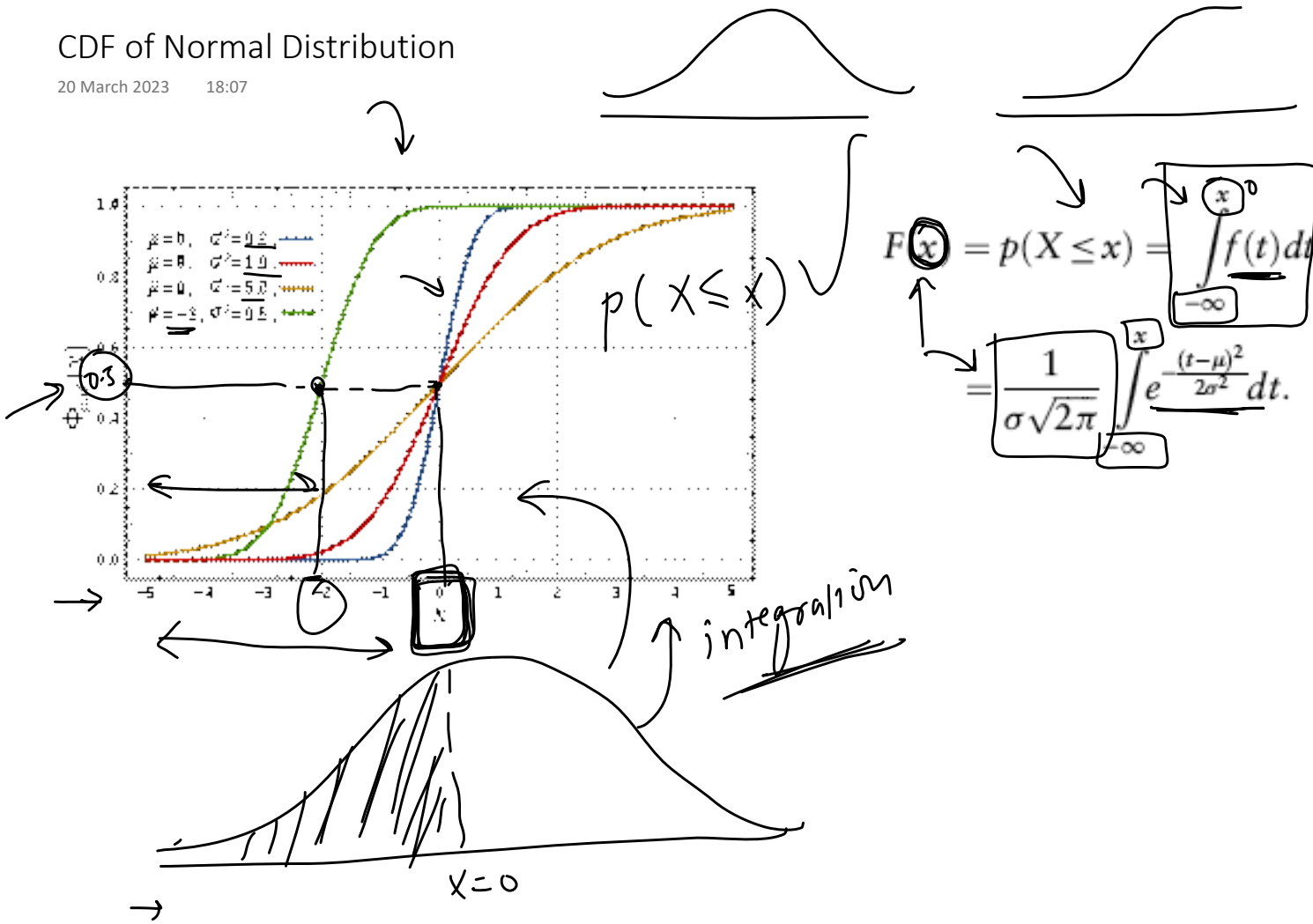
Python Example

Interpretation



CDF of Normal Distribution



20 March 2023 18:07



Use in Data Science

20 March 2023

18:08

- 
- 
- Outlier detection
 - Assumptions on data for ML algorithms -> Linear Regression and GMM
 - Hypothesis Testing
 - Central Limit Theorem