Train In Data

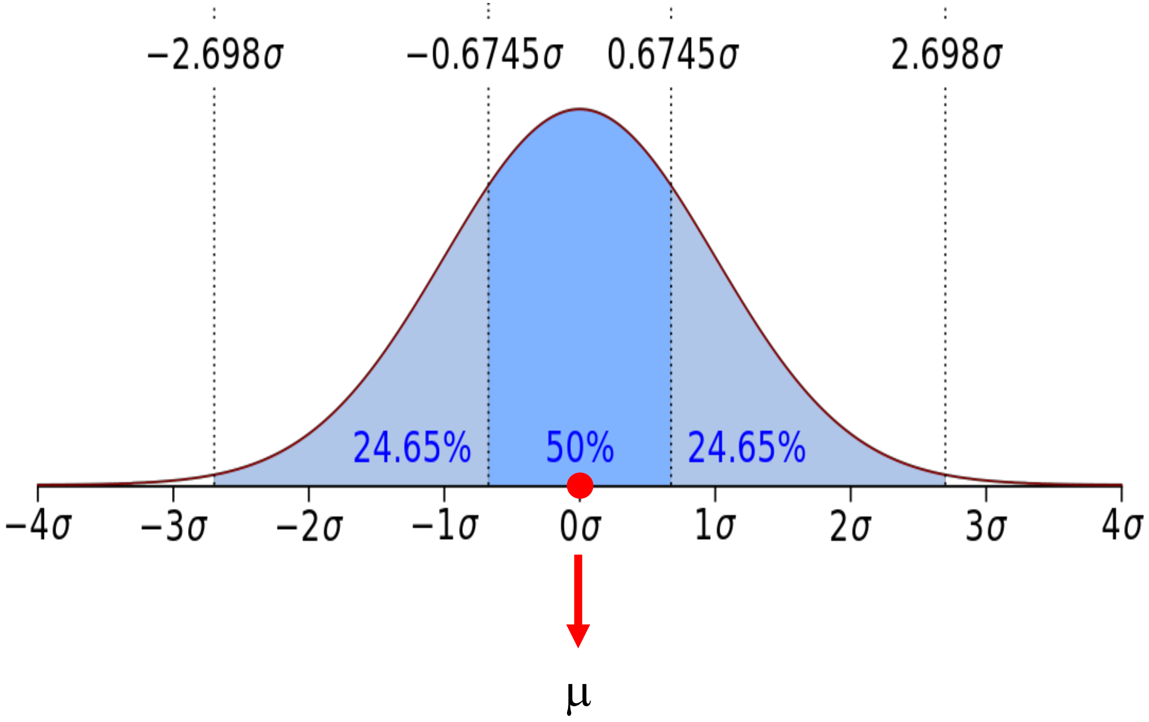
Gaussian Distribution

Gaussian distribution

We say a variable follows Gaussian dist. when it adopts this bell shape. The bell shape indicates the prob. of finding an observation with a certain value in the dist. of this variable.

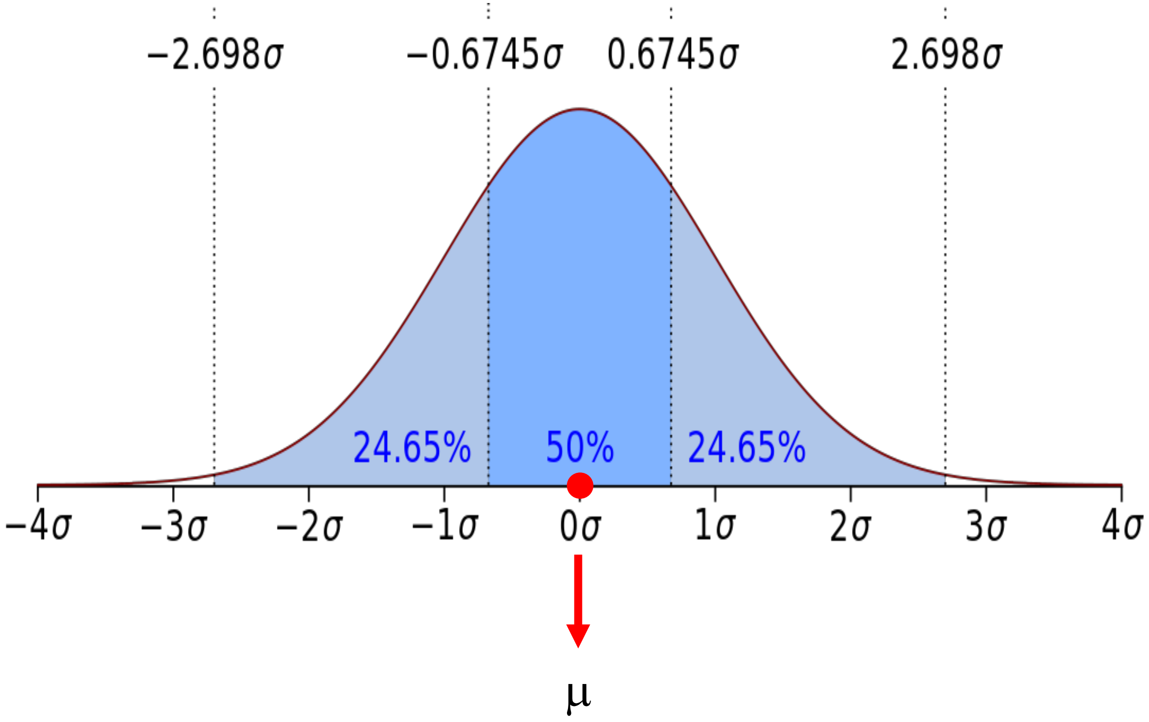
* This bell is centered around the mean (μ) value of the var.

* The Gaussian dist. is also characterized by std. dev. (σ) => how dispersed the values are around the mean.



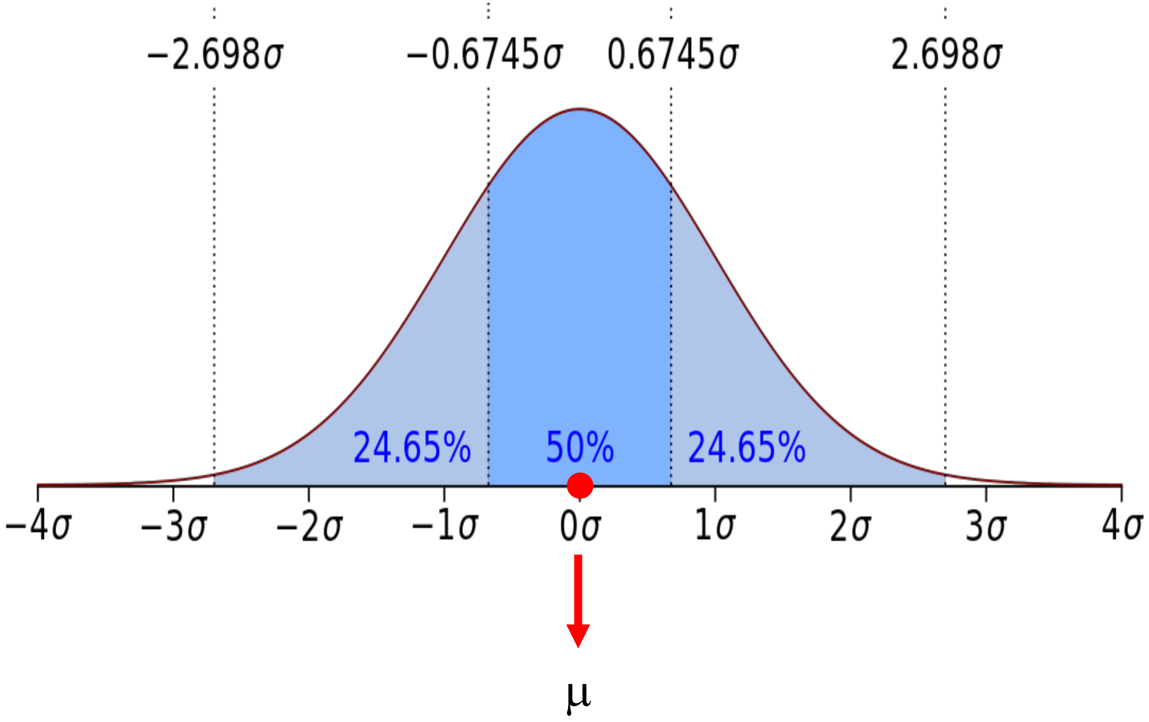
- Bell shape
- μ = Mean value
  - ➔ centre of distribution
- σ = standard deviation
  - ➔ measure of dispersion

Gaussian distribution



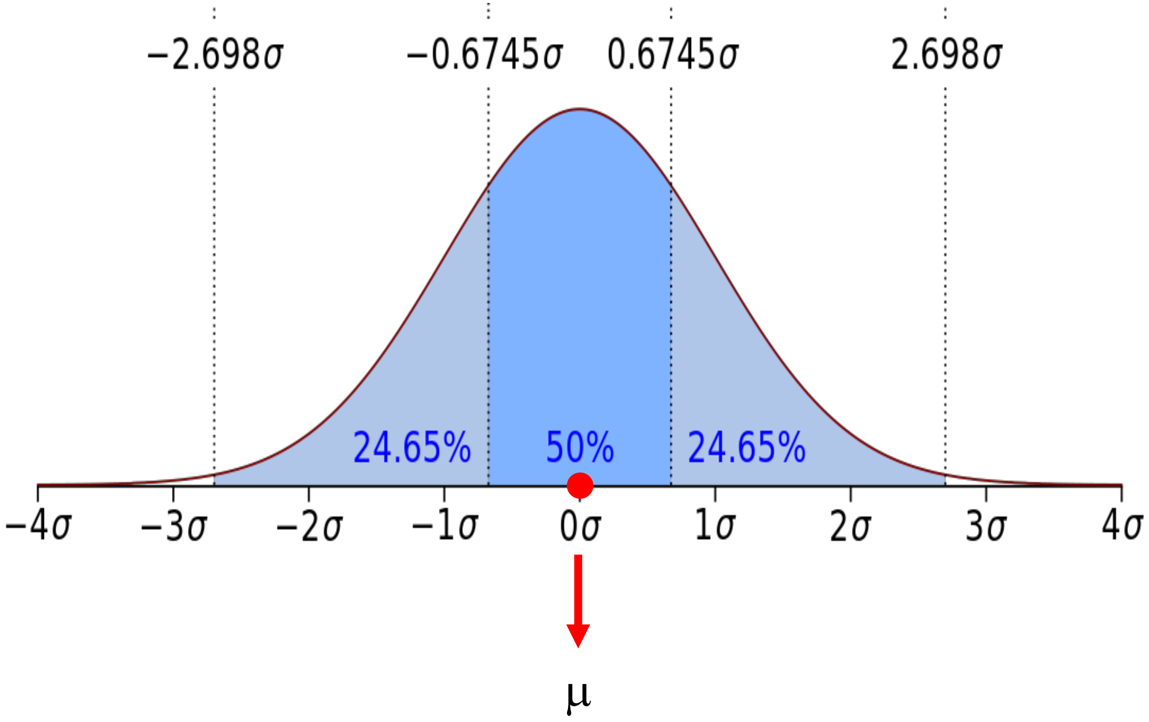
- **Symmetric:**
  - Most observations occur around the centre
  - Probabilities for values further away from the centre decrease equally in both directions.
  - Extreme values in both tails of the distribution are similarly unlikely.

Gaussian distribution



- ~50% of the observations within $x_{mean} \pm 0.67 \times \sigma$.
- ~99% of the observations within $x_{mean} \pm 2.7 \times \sigma$.

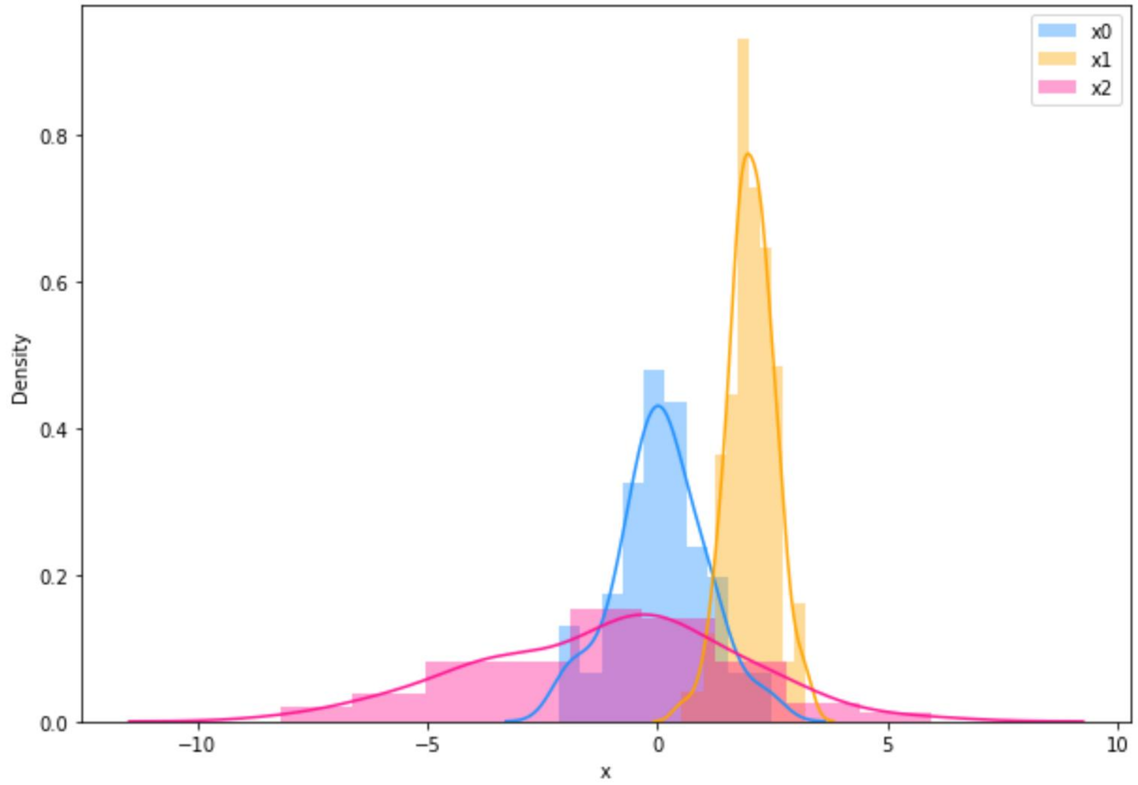
Gaussian distribution



A variable is normally distributed:

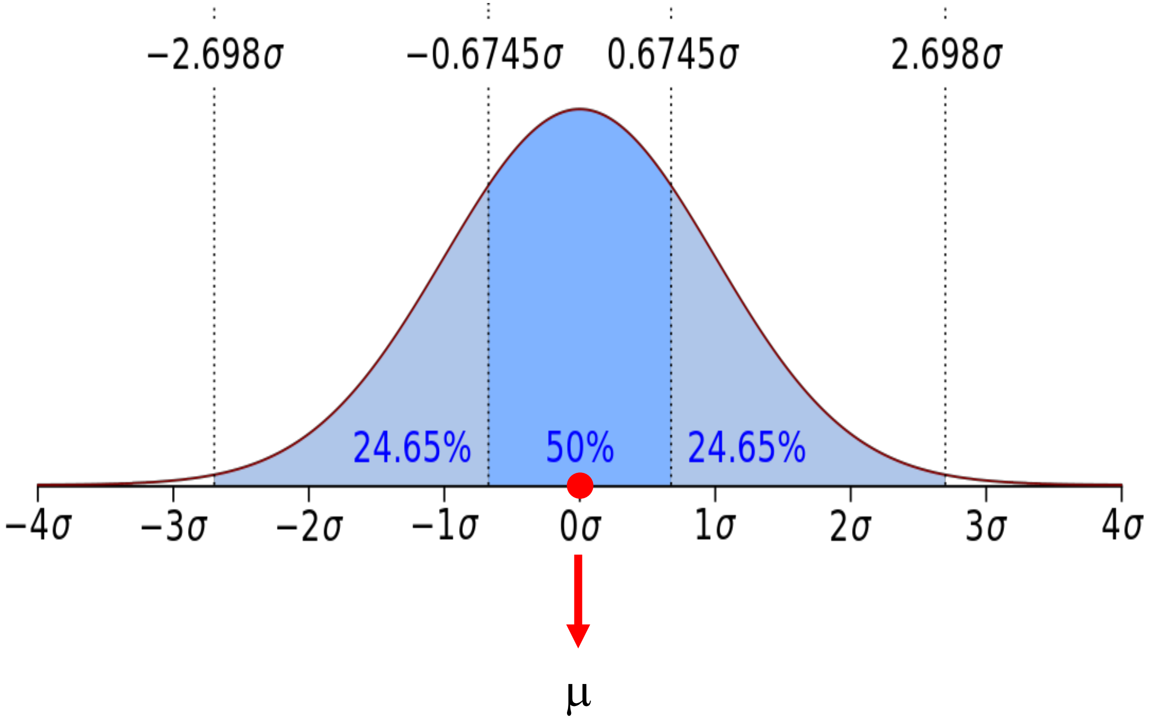
- $X \sim N(\mu, \sigma^2)$
- $X1 \sim N(\mu=0, \sigma^2=1)$

Gaussian distribution



- $X0 \sim N(\mu = 0,\ \sigma^2 = 1)$
- $X1 \sim N(\mu = 2,\ \sigma^2 = 0.5)$
- $X2 \sim N(\mu = -1,\ \sigma^2 = 3)$

Gaussian distribution



$X \sim N(\mu, \sigma^2)$

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

Mean

- Average of the variable values

- $X_{mean} = \mu = \frac{\sum x_i}{n}$

- n= number of observations

Mean

- Average of the variable values
- $X_{mean} = \mu = \frac{\sum x_i}{n}$
- n= number of observations

X1	X2
12	10
8	7
16	13

- $\mu1 = (12 + 8 + 16) / 3 = 12$
- $\mu2 = (10 + 7 + 13) / 3 = 10$

Variance and Standard Deviation

- Measure the dispersion of the data, away from the mean
- var = $\sigma^2 = \frac{\sum(x_i - x_{mean})^2}{n}$
- n= number of observations
- σ = standard deviation.
- $\sigma = \sqrt[2]{var}$

Variance and Standard Deviation

- Measure the dispersion of the data, away from the mean
- var = $\frac{\sum(x_i - x_{mean})^2}{n-1}$
- n= number of observations
- σ = standard deviation.
- σ = $\sqrt[2]{var}$

X1	X2
12	10
8	7
16	13

- σ1 = $\sqrt[2]{\frac{(12-12)^2 + (8-12)^2 + (16-12)^2}{3-1}}$ = 4
- σ2 = $\sqrt[2]{\frac{(10-10)^2 + (7-10)^2 + (13-10)^2}{3-1}}$ = 3

THANK YOU

www.trainindata.com