✨ Why Gaussian Distribution?

When a feature (like Age or Income) is **continuous**, Naïve Bayes assumes the values are **normally distributed**. So we calculate the **likelihood P(x | Class)** using the **Gaussian Probability Density Function (PDF):

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

🔢 Step 1: Sample Training Data

ID	Age	Income (₹k)	Buys_Computer
1	25	40	Yes
2	30	42	Yes
3	35	45	Yes
4	40	70	No
5	45	65	No
6	50	80	No

🎯 Step 2: Input Tuple to Classify

Let `X = (Age = 38, Income = 50)`

📊 Step 3: Calculate Mean (μ) and Standard Deviation (σ)

For class "Yes" (Tuples 1, 2, 3):

Age:

25+30+35

$$\mu_{\text{Yes, Age}} = \frac{\cdots}{3} = 30$$

$$\sigma_{\text{Yes, Age}} = \sqrt{\frac{(25-30)^2 + (30-30)^2 + (35-30)^2}{3}} = \sqrt{\frac{25+0+25}{3}} = \sqrt{16.67} \approx 4.08$$

Income:

$$\mu_{\text{Yes, Income}} = \frac{40+42+45}{3} = \frac{127}{3} \approx 42.33$$

$$\sigma_{\text{Yes, Income}} = \sqrt{\frac{(40-42.33)^2 + (42-42.33)^2 + (45-42.33)^2}{3}}$$

$$= \sqrt{\frac{(-2.33)^2 + (-0.33)^2 + (2.67)^2}{3}} = \sqrt{\frac{5.43+0.11+7.13}{3}} = \sqrt{4.22} \approx 2.05$$

For class "No" (Tuples 4, 5, 6):

Age:

$$\mu_{\text{No, Age}} = \frac{40+45+50}{3} = 45$$

$$\sigma_{\text{No, Age}} = \sqrt{\frac{(40-45)^2 + (45-45)^2 + (50-45)^2}{3}} = \sqrt{\frac{25+0+25}{3}} = \sqrt{16.67} \approx 4.08$$

Income:

$$\mu_{\text{No, Income}} = \frac{70+65+80}{3} = \frac{215}{3} \approx 71.67$$

$$\sigma_{\text{No, Income}} = \sqrt{\frac{(70-71.67)^2 + (65-71.67)^2 + (80-71.67)^2}{3}}$$

$$= \sqrt{\frac{(-1.67)^2 + (-6.67)^2 + (8.33)^2}{3}} = \sqrt{\frac{2.79+44.49+69.39}{3}} = \sqrt{38.89} \approx 6.24$$

🧮 Step 4: Use Gaussian PDF to Compute Likelihoods

The formula:

$$P(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

For Class = Yes:

1️⃣ **P(Age = 38 | Yes)**

$$\mu = 30,\ \sigma = 4.08$$

$$P = \frac{1}{\sqrt{2\pi} \cdot 4.08} \cdot e^{-\frac{(38-30)^2}{2\cdot(4.08)^2}} = \frac{1}{10.22} \cdot e^{-\frac{64}{33.3}} \approx 0.014$$

2️⃣ **P(Income = 50 | Yes)**

$$\mu = 42.33,\ \sigma = 2.05$$

$$P = \frac{1}{\sqrt{2\pi} \cdot 2.05} \cdot e^{-\frac{(50-42.33)^2}{2\cdot(2.05)^2}} = \frac{1}{5.13} \cdot e^{-\frac{58.78}{8.4}} \approx 1.95 \times 10^{-5}$$

$$P(X \mid \text{Yes}) = 0.014 \cdot 1.95 \times 10^{-5} = 2.73 \times 10^{-7}$$

For Class = No:

1️⃣ **P(Age = 38 | No)**

$$\mu = 45,\ \sigma = 4.08$$

$$P = \frac{1}{\sqrt{2\pi} \cdot 4.08} \cdot e^{-\frac{(38-45)^2}{2\cdot(4.08)^2}} = \frac{1}{10.22} \cdot e^{-\frac{49}{33.3}} \approx 0.044$$

2️⃣ **P(Income = 50 | No)**

$$\mu = 71.67,\ \sigma = 6.24$$

$$P = \frac{1}{\sqrt{2\pi} \cdot 6.24} \cdot e^{-\frac{(50-71.67)^2}{2\cdot(6.24)^2}} = \frac{1}{15.63} \cdot e^{-\frac{475.69}{77.88}} \approx 1.75 \times 10^{-4}$$

$$P(X \mid \text{No}) = 0.044 \cdot 1.75 \times 10^{-4} = 7.7 \times 10^{-6}$$

🔢 Step 5: Prior Probabilities

⚖️ Step 6: Posterior Probabilities

$$P(\text{Yes} \mid X) \propto P(X \mid \text{Yes}) \cdot P(\text{Yes}) = 2.73 \times 10^{-7} \cdot 0.5 = 1.36 \times 10^{-7}$$

$$P(\text{No} \mid X) \propto P(X \mid \text{No}) \cdot P(\text{No}) = 7.7 \times 10^{-6} \cdot 0.5 = 3.85 \times 10^{-6}$$

✅ Final Prediction

Since

$$P(\text{No} \mid X) > P(\text{Yes} \mid X)$$

🔮 The classifier **predicts "No"** (i.e., does **not** buy a computer).

📌 Summary: Why Gaussian Distribution Matters

- For continuous features, exact probability counts don't work.
- Gaussian distribution allows Naïve Bayes to estimate probability using **mean and variance**.
- It's simple, efficient, and surprisingly effective when the normality assumption roughly holds.
