Bayes' Theorem

The correct form of Bayes' Theorem for classification is:

$$P(\text{Yes} \mid \text{Features}) = \frac{P(\text{Age} \mid \text{Yes}) \cdot P(\text{Sex} \mid \text{Yes}) \cdot P(\text{Chest Pain} \mid \text{Yes}) \cdot P(\text{Yes})}{P(\text{Age}) \cdot P(\text{Sex}) \cdot P(\text{Chest Pain})}$$

Where:

- **P(Yes | Features)**: The probability of having heart disease given the features (Age, Sex, Chest Pain).
- **P(Age | Yes)**: The likelihood of observing **Age = 50** for someone with heart disease.
- **P(Sex | Yes)**: The likelihood of observing **Sex = Male** for someone with heart disease.
- **P(Chest Pain | Yes)**: The likelihood of observing **Chest Pain = Typical Angina** for someone with heart disease.
- **P(Yes)**: The prior probability of having heart disease.
- **P(Age)**, **P(Sex)**, **P(Chest Pain)**: The marginal probabilities of observing **Age = 50**, **Sex = Male**, and **Chest Pain = Typical Angina** in the dataset (irrespective of having heart disease).

Let's calculate each part step-by-step, using the dataset provided:

Age	Sex	Chest Pain Type	Target (Heart Disease)
45	Male	Typical Angina	Yes
54	Female	Atypical Angina	No
60	Male	Non-Anginal Pain	Yes
50	Male	Atypical Angina	No

Step 1: Calculate **Prior Probability** (P(Yes))

From the dataset, we know that there are **2 instances with heart disease** (Yes) and **4 total instances**:

$$P(\text{Yes}) = \frac{\text{Number of Yes cases}}{\text{Total number of cases}} = \frac{2}{4} = 0.5$$

Step 2: Calculate **Likelihoods** (P(Feature | Yes))

1. **P(Age | Yes)**:
    - We have **2 instances with Heart Disease = Yes**. The ages are **45** and **60**.
    - **Age = 50** is between these two values, so we can compute the **Gaussian distribution** for this continuous feature.

    The mean (μ) of ages for **Yes** is:

$$\mu_{\text{Yes}} = \frac{45 + 60}{2} = 52.5$$

The variance (σ^2) for ages of **Yes** is:

$$\sigma^2_{\text{Yes}} = \frac{(45 - 52.5)^2 + (60 - 52.5)^2}{2} = \frac{56.25 + 56.25}{2} = 56.25$$

Now, using the **Gaussian distribution**:

$$P(\text{Age} = 50 \mid \text{Yes}) = \frac{1}{\sqrt{2\pi \cdot 56.25}} \cdot e^{-\frac{(50-52.5)^2}{2 \cdot 56.25}} \approx 0.037$$

---

2. **P(Sex | Yes)**:
   - There are **2 males with heart disease** (45 and 60).
   - So:

$$P(\text{Male} \mid \text{Yes}) = \frac{2}{2} = 1.0$$

---

3. **P(Chest Pain | Yes)**:
   - There is **1 instance of Typical Angina** in the "Yes" class (Age 45).
   - So:

$$P(\text{Typical Angina} \mid \text{Yes}) = \frac{1}{2} = 0.5$$

Step 3: Calculate **Marginal Probabilities** (P(Feature))

These are the probabilities of observing each feature in the dataset, irrespective of whether the target is **Yes** or **No**.

1. **P(Age)**:
   - The ages in the dataset are: **45, 54, 60, 50**.
   - The probability of **Age = 50** is calculated using the Gaussian distribution for all ages, which are distributed with a mean of **52.25** and variance calculated from all 4 instances.

   The **mean** and **variance** for the dataset:

$$\mu_{\text{All}} = \frac{45 + 54 + 60 + 50}{4} = 52.25$$

$$\sigma_{\text{All}}^2 = \frac{(45-52.25)^2 + (54-52.25)^2 + (60-52.25)^2 + (50-52.25)^2}{4} = 34.1875$$

Now, calculate:

$$P(\text{Age}=50) = \frac{1}{\sqrt{2\pi \cdot 34.1875}} \cdot e^{-\frac{(50-52.25)^2}{2\cdot 34.1875}} \approx 0.099$$

---

2. **P(Sex)**:
   - There are **3 males** and **1 female** in the dataset.

$$P(\text{Male}) = \frac{3}{4} = 0.75$$

---

3. **P(Chest Pain)**:
   - The chest pain types in the dataset are: **Typical Angina** (2 instances), **Atypical Angina** (2 instances), **Non-Anginal Pain** (1 instance).
   - So:

$$P(\text{Typical Angina}) = \frac{2}{5} = 0.4$$

Step 4: Calculate Posterior Probability for Yes

Now we can substitute the values into the formula for **P(Yes | Features)**:

$$P(\text{Yes} \mid \text{Features}) = \frac{P(\text{Age}=50 \mid \text{Yes}) \cdot P(\text{Male} \mid \text{Yes}) \cdot P(\text{Typical Angina} \mid \text{Yes}) \cdot P(\text{Yes})}{P(\text{Age}) \cdot P(\text{Male}) \cdot P(\text{Typical Angina})}$$

Substitute the values:

$$P(\text{Yes} \mid \text{Features}) = \frac{0.037 \cdot 1.0 \cdot 0.5 \cdot 0.5}{0.099 \cdot 0.75 \cdot 0.4} \approx \frac{0.00925}{0.0297} \approx 0.311$$

Step 5: Calculate Posterior Probability for No

Similarly, we calculate the posterior probability for **No**:

$$P(\text{No} \mid \text{Features}) = \frac{P(\text{Age}=50 \mid \text{No}) \cdot P(\text{Male} \mid \text{No}) \cdot P(\text{Typical Angina} \mid \text{No}) \cdot P(\text{No})}{P(\text{Age}) \cdot P(\text{Male}) \cdot P(\text{Typical Angina})}$$

We already know the marginal probabilities and likelihoods for **No**:

- **P(Age = 50 | No)** = 0.2419
- **P(Male | No)** = 1.0
- **P(Typical Angina | No)** = 0.0
- **P(No)** = 0.5

Substitute the values:

$$P(\text{No} \mid \text{Features}) = \frac{0.2419 \cdot 1.0 \cdot 0.0 \cdot 0.5}{0.099 \cdot 0.75 \cdot 0.4} = 0$$

Step 6: Decision

Now we compare the two posterior probabilities:

$$P(\text{Yes} \mid \text{Features}) = 0.311$$

$$P(\text{No} \mid \text{Features}) = 0$$

Since **P(Yes | Features)** > **P(No | Features)**,