Email	Contains "buy"	Contains "cheap"	Contains "click"	Spam
1	Yes	Yes	No	Yes
2	Yes	No	Yes	Yes
3	No	Yes	No	No
4	No	Yes	Yes	No
5	Yes	No	Yes	Yes

Now, we want to classify a new email that contains "buy" and "click" but not "cheap". Let's denote:

- C_1 = Spam
- C_2 = Not Spam
- x_1 = Contains "buy" = Yes
- x_2 = Contains "cheap" = No
- x_3 = Contains "click" = Yes

We need to compute $P(C_1|x_1, x_2, x_3)$ and $P(C_2|x_1, x_2, x_3)$ and then classify the email based on which probability is higher.

Steps:

1. **Calculate Priors:**

$$P(C_1) = \frac{\text{Number of Spam emails}}{\text{Total emails}} = \frac{3}{5} = 0.6$$

$$P(C_2) = \frac{\text{Number of Not Spam emails}}{\text{Total emails}} = \frac{2}{5} = 0.4$$

2. **Calculate Likelihoods:**

For $P(x_1 = \text{Yes}|C_1)$:

$$P(x_1 = \text{Yes}|C_1) = \frac{\text{Number of Spam emails with "buy"}}{\text{Total Spam emails}} = \frac{2}{3} \approx 0.67$$

For $P(x_2 = \text{No}|C_1)$:

$$P(x_2 = \text{No}|C_1) = \frac{\text{Number of Spam emails without "cheap"}}{\text{Total Spam emails}} = \frac{1}{3} \approx 0.33$$

For $P(x_3 = \text{Yes}|C_1)$:

$$P(x_3 = \text{Yes}|C_1) = \frac{\text{Number of Spam emails with "click"}}{\text{Total Spam emails}} = \frac{2}{3} \approx 0.67$$

For $P(x_1 = \text{Yes}|C_2)$:

$$P(x_1 = \text{Yes}|C_2) = \frac{\text{Number of Not Spam emails with "buy"}}{\text{Total Not Spam emails}} = \frac{0}{2} = 0$$

For $P(x_2 = \text{No}|C_2)$:

$$P(x_2 = \text{No}|C_2) = \frac{\text{Number of Not Spam emails without "cheap"}}{\text{Total Not Spam emails}} = \frac{1}{2} = 0.5$$

For $P(x_3 = \text{Yes}|C_2)$:

$$P(x_3 = \text{Yes}|C_2) = \frac{\text{Number of Not Spam emails with "click"}}{\text{Total Not Spam emails}} = \frac{1}{2} = 0.5$$

Dataset

Let's use the same dataset as before:

Email	Contains "buy"	Contains "cheap"	Contains "click"	Spam
1	Yes	Yes	No	Yes
2	Yes	No	Yes	Yes
3	No	Yes	No	No
4	No	Yes	Yes	No
5	Yes	No	Yes	Yes

Step-by-Step Process

1. **Feature Representation:**

   We'll represent the features as binary values (1 for Yes, 0 for No):

Contains "buy"	Contains "cheap"	Contains "click"	Spam
1	1	0	1
1	0	1	1
0	1	0	0
0	1	1	0
1	0	1	1

2. **Separate Features and Labels:**

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$y = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

3. **Train the SVM:**

   We use a linear SVM to train on this dataset. The objective is to find the hyperplane that best separates the spam (1) from the non-spam (0) emails.

   The SVM will solve the optimization problem to find the weight vector $\mathbf{w}$ and bias b that define the hyperplane.

3. **Train the SVM:**

   We use a linear SVM to train on this dataset. The objective is to find the hyperplane that best separates the spam (1) from the non-spam (0) emails.

   The SVM will solve the optimization problem to find the weight vector $\mathbf{w}$ and bias b that define the hyperplane.

4. **Mathematical Formulation:**

   The decision function for the SVM classifier is:

   $$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$$

   Where:

   - $\mathbf{w}$ is the weight vector
   - $\mathbf{x}$ is the feature vector
   - b is the bias term

   The predicted class is given by the sign of $f(\mathbf{x})$:

   $$\text{Class} = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$$

5. **Example Classification:**

   Let's classify a new email with the following features:

   - Contains "buy": Yes (1)
   - Contains "cheap": No (0)
   - Contains "click": Yes (1)

   Feature vector:

$$\mathbf{x}_{\text{new}} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

   Assuming after training we obtained $\mathbf{w} = [2, -1, 1]$ and $b = -1$, we compute:

$$f(\mathbf{x}_{\text{new}}) = [2, -1, 1] \cdot [1, 0, 1] + (-1)$$

$$f(\mathbf{x}_{\text{new}}) = (2 \times 1) + (-1 \times 0) + (1 \times 1) + (-1) = 2 + 0 + 1 - 1 = 2$$

   The predicted class is:

$$\text{Class} = \text{sign}(2) = 1 \quad \text{(spam)}$$