1. **What is marginal probability**

It is a probability of one event occurred calculation by ignoring all other variables.





1. **What are the probability axioms?**

It is called rules or principles of probability theory, has three rules





This calculation will used in **Naive Bayes** where naïve mean assumption, so in **Naive Bayes** is a **probabilistic classifier** based on **Bayes' Theorem** that assumes features are independent — it predicts the class with the highest probability.

1. **What is Conditional probability ?**

Conditional probability is the chance of event **A** happening **if** event **B** has already happened — like the chance it’s raining (**A**) given that it’s cloudy (**B**).

**Example**: If 30 out of 50 cloudy days are rainy, then  
**P(Rain | Cloudy) = 30 / 50 = 0.6**.

1. **What is Bayes’ Theorem and when is it used in data science?**

Bayes' Theorem helps us find the probability of a cause given an observed result. For example, if a person tests positive for a disease, Bayes' Theorem helps calculate how likely they truly have it, considering the test’s accuracy and disease rate. It's useful in areas like medical diagnosis, spam detection, and predictive modeling.

 Bayes' Theorem = Formula for conditional probability

 Naive Bayes = A classification model built using Bayes’ Theorem

Naive Bayes is widely used for spam detection, sentiment analysis, and text classification.

1. **Define variance and conditional variance.**

**Variance**: It shows how spread out numbers are from the average — for example, in the numbers [2, 4, 6], the variance is small because they’re close to the mean (4).

**Conditional Variance**: It’s the variance of a variable given some condition — for example, if we know it’s a rainy day, the variance of people carrying umbrellas may be less than on random days.

**EX**:

**🔹 First, What is Conditional Variance?**

The **conditional variance** of a random variable YYY given another variable XXX is:

Var(Y∣X)=E[(Y−E[Y∣X])2∣X]\text{Var}(Y \mid X) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X]Var(Y∣X)=E[(Y−E[Y∣X])2∣X]

It measures how much **Y varies**, *given the value of* XXX.

**🔸 Simple Example (Discrete)**

Let’s say you have the following data showing a variable XXX (like a group or category), and a response variable YYY (like test scores):

| **X (Group)** | **Y (Score)** |
| --- | --- |
| A | 70 |
| A | 75 |
| A | 65 |
| B | 90 |
| B | 95 |
| B | 85 |

Let’s compute the **conditional variance of Y given X=A** and **X=B**.

**📌 Step-by-step for X = A:**

Y values when X=A: [70, 75, 65]

* Mean of Y|X=A = (70+75+65)/3=70(70 + 75 + 65)/3 = 70(70+75+65)/3=70
* Variance of Y|X=A:

(70−70)2+(75−70)2+(65−70)23=0+25+253=503≈16.67\frac{(70 - 70)^2 + (75 - 70)^2 + (65 - 70)^2}{3} = \frac{0 + 25 + 25}{3} = \frac{50}{3} \approx 16.673(70−70)2+(75−70)2+(65−70)2​=30+25+25​=350​≈16.67

**📌 For X = B:**

Y values when X=B: [90, 95, 85]

* Mean of Y|X=B = (90+95+85)/3=90(90 + 95 + 85)/3 = 90(90+95+85)/3=90
* Variance of Y|X=B:

(90−90)2+(95−90)2+(85−90)23=0+25+253=503≈16.67\frac{(90 - 90)^2 + (95 - 90)^2 + (85 - 90)^2}{3} = \frac{0 + 25 + 25}{3} = \frac{50}{3} \approx 16.673(90−90)2+(95−90)2+(85−90)2​=30+25+25​=350​≈16.67

So in this example:

* Var(Y∣X=A)≈16.67\text{Var}(Y \mid X=A) \approx 16.67Var(Y∣X=A)≈16.67
* Var(Y∣X=B)≈16.67\text{Var}(Y \mid X=B) \approx 16.67Var(Y∣X=B)≈16.67

1. **Explain the concepts of mean, median, mode, and standard deviation.**

**Mean:**The mean, often referred to as the average, is calculated by summing up all the values in a dataset and then dividing by the total number of values.

**Median:**When data are sorted in either ascending or descending order, the median is the value in the middle of the dataset. The median is the average of the two middle values when the number of data points is even. In comparison to the mean, the median is less impacted by extreme numbers, making it a more reliable indicator of central tendency.

**Mode:** The value that appears most frequently in a dataset is the mode. One mode (unimodal), several modes (multimodal), or no mode (if all values occur with the same frequency) can all exist in a dataset.

**Standard deviation:** The spread or dispersion of data points in a dataset is measured by the standard deviation. It quantifies the variance between different data points