Bayes' theorem

Let A and B two events (**Outcomes**) , Probability of A given B happened

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Called **Conditional Probability**
- **Law of Total Probability** $\rightarrow P(A \cap B) = P(A|B)P(B)$
- **Given** B means that the event B already happened
  The **Bayes** Theorem:
- $P(B|A) = ?$ - B is an event **hard to measure** (expansive and cost efforts and money)
- A is an event **easy to measure** (Cheap and easy)

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- $P(B|A) \rightarrow$ Posterior
- $P(B) \rightarrow$ Prior
- $P(A|B) \rightarrow$ Update

$$\text{Prior: } P(B) \quad \xrightarrow{\text{observe } A} \quad \text{Posterior: } P(B|A) \quad \Rightarrow \text{New Prior for next update}$$

We use $P(B)$ as a base information and we use " The **easy to measure** $P(A|B)$" to **update** $P(B) \rightarrow P(B|A)$.
So the more information we gather $P(B|A) \rightarrow$ will replace $P(B)$ as the base information

Without $P(A)$:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

- $P(A|B)P(B) \rightarrow P(A \cap B)$
- $P(A|B^c)P(B^c) \rightarrow P(A \cap B^c)$
- $P(A \cap B) + P(A|B^c)P(B^c) = P(A)$
  Generalized Byes' Theorem :

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum P(A|B_j)P(B_j)}$$

- Assuming B_j are **Disjoint** events $B_i \cap B_j = \emptyset$

Conclusion

- Bayes' rule forms the foundation of Generative Models for Classification Like **LDA** which estimate $P(X|Y)$ and $P(Y)$, then apply Bayes' Theorem to compute $P(Y|X)$