Naïve Bayes

- Probabilistic classification algorithm based on Bayes' theorem

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

$P(A|B)$: Probability of A given B

$P(A)$: Probability of A

- Recap of probability
  - Probability: Likelihood of an even occurring $P(A) \longrightarrow 0 \leq P(A) \leq 1$
  - Complement Rule: $\neg P(A) = 1 - P(A)$
  - Mutually exclusive events (No overlap): Both events cannot happen at the same time $P(A \cap B) = 0$

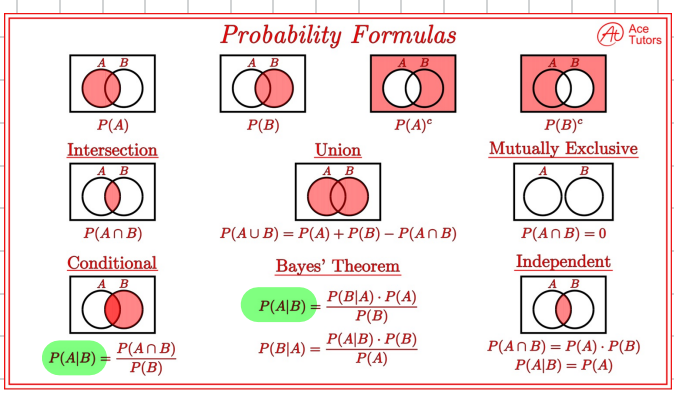
    Rule for mutually exclusive events: $P(A \cup B) = P(A) + P(B)$

    If two events are mutually exclusive, they are always dependent (one happening means the other cannot
  - Overlapping: Both events can happen at the same time $P(A \cap B) \neq 0$

    Rule for overlapping events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Independent events: The occurrance of one does not affect the probability of the other $P(A|B) = P(A)$

    Independent: we use $P(A \cap B) = P(A) \cdot P(B)$
    - Filling a coin and rolling a die
    - Rolling two dice
  - Dependent events: The occurrance of one affect the probability of the other $P(A|B) \neq P(A)$

    Dependent: we use $P(A|B) = \frac{P(A \cap B)}{P(B)}$
    - Studying and pass an exam
    - It rains and take an umbrella



Mutually exclusive are dependent events

Overlapping can be dependent or independent

- **Distribution:** Describes how values are spread or dispersed. Distributions can be discrete or continuous (infinite values within a given range)

- **Distributions**

  - **Normal (Gaussian) Distribution:** Symmetric around its mean. Characterized by its mean and distribution. It's a continuous distribution. Example: Heights, IQ scores

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

  - **Binomial distribution:** Multiple trials, two only possible outcomes. It's a discrete distribution. Example: Get 10 times head flipping a coin

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

n: number of trials
p: probability of success
k: number of success

  - **Bernoulli distribution:** One trial, two only possible outcomes. It's discrete. Example: Flip a coin

$$P(X=x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}$$

  - **Categorical distribution:** Generalizes Bernoulli distribution. Here, one trial, multiple options. Example: Rolling a die once

$$P(X=x_k) = p_k, \quad \sum_{k=1}^{K} p_k = 1$$

p_k: probability of category k

  - **Multinomial distribution:** Generalizes Binomial distribution: Here, multiple trials, multiple options. Example: Rolling a die 10 times

$$P(X_1=x_1, \ldots, X_k=x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdots p_k^{x_k}$$

n: total of trials
x_k: number of times of k
p_k: probability of category k
$\sum x_k = n$ (all trials must sum to n)

- Naive Bayes assumes all variables are independent. This means:

$$P(X_1, X_2, X_3 \ldots X_n | Y) = P(X_1|Y) \cdot P(X_2|Y) \cdot \ldots \cdot P(X_n|Y)$$

- We got three types of Naive Bayes:

  - **Gaussian Naive Bayes (for continuous data)**

$$P(X_i|Y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(X_i-\mu_y)^2}{2\sigma_y^2}}$$

- Assumes features follow normal (Gaussian) distribution

  - **Multinomial Naive Bayes (for text data)**

$$P(X_i|Y) = \frac{\text{count}(X_i|Y) + \alpha}{\sum_j \text{count}(X_j|Y) + \alpha \cdot V}$$

- Discrete features
- $\text{count}(X_i|Y)$: times X_i appear in class Y
- V: Vocabulary size
- α: Smoothing parameter

  - **Bernoulli Naive Bayes (for binary features)**

$$P(X_i|Y) = p^{x_i}(1-p)^{(1-x_i)}$$

Example Gaussian Nayes

Feature 1	Feature 2	Class (Y)
6	180	0 (Male)
5.5	160	0 (Male)
5.8	170	0 (Male)
5.2	140	1 (Female)
5	130	1 (Female)
5.3	135	1 (Female)

Data to classify: $X = (5.7, 165)$

1. Calculate the mean and variance of each class

Class 0

mean

$$\mu_{x_1,0} = \frac{6+5.5+5.8}{3} = \frac{17.3}{3} = 5.77$$

$$\mu_{x_2,0} = \frac{180+160+170}{3} = \frac{510}{3} = 170$$

variance

$$\sigma^2_{x_1,0} = \frac{(6-5.7)^2+(5.5-5.7)^2+(5.8-5.7)^2}{3} = 0.058$$

$$\sigma^2_{x_2,0} = \frac{(180-170)^2+(160-170)^2+(160-170)^2}{3} = 66.67$$

Class 1

mean

$$\mu_{x_1,1} = \frac{5.2+5+5.3}{3} = \frac{15.5}{3} = 5.17$$

$$\mu_{x_2,1} = \frac{140+130+135}{3} = \frac{405}{3} = 135$$

variance

$$\sigma_{x_1,1} = \frac{(5.2-5.17)^2+(5-5.17)^2+(5.3-5.17)^2}{3} = 0.017$$

$$\sigma_{x_2,1} = \frac{(140-135)^2+(130-135)^2+(135-135)^2}{3} = 16.67$$

2. Compute Gaussian probability

$$P(X_i|C_k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)}$$

Class 0

$$P(5.7|0) = \frac{1}{\sqrt{2\cdot\pi\cdot 0.058}} e^{\left(-\frac{(5.7-5.77)^2}{2\cdot 0.058}\right)} = 1.57$$

$$P(165|0) = \frac{1}{\sqrt{2\cdot\pi\cdot 66.67}} e^{\left(-\frac{(165-170)^2}{2\cdot 66.67}\right)} = 0.046$$

Class 1

$$P(5.7|1) = \frac{1}{\sqrt{2\cdot\pi\cdot 0.017}} e^{-\frac{(5.7-5.17)^2}{2\cdot 0.017}} = 0.0$$

$$P(165|1) = \frac{1}{\sqrt{2\cdot\pi\cdot 16.67}} e^{-\frac{(165-135)^2}{2\cdot 16.67}} = 0.0$$

$\Rightarrow X = (5.7, 165)$ corresponds to 0 (male)