NAïVE BAYES

What is Naïve Bayes ?

Naïve Bayes Theorem หรือ ตัวจำแนกแบบเบย์อย่างง่าย (Naive Bayesian Classifier) คือโมเดลการจำแนกประเภทข้อมูล ที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem และสมมติฐาน ที่ให้การเกิดของเหตุการณ์ต่าง ๆ ซึ่งเป็นอิสระต่อกัน (Independence) แบบมีเงื่อนไข (Conditional) หรือมีความเป็นอิสระต่อกันแบบมีเงื่อนไข (Conditional Independence)

Conditional Independence

ความเป็นอิสระต่อกันแบบมีเงื่อนไขเป็นการช่วยให้ระบบความน่าจะเป็นมีขนาดเพียงแค่ $O(n)$ แทนที่จะเป็น $O(2^n)$ เพื่อทำให้ค่าใกล้เคียงกับความเป็นจริงให้มากที่สุด

โดย Conditional Independence มาจาก 2 ทฤษฎี คือ 1.Conditional และ 2.Independence

โดยสามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$P(X,Y|Z) = P(X|Z)\,P(Y|Z)$$

$P(X|Y,Z) = P(X|Z)$ และ $P(Y|X,Z) = P(Y|Z)$

ซึ่งตัวแปร X และ Y จะเป็นเหตุการณ์ที่เป็นอิสระต่อกันแต่มีตัวแปร Z มาเป็นเงื่อนไข

Conditional

เป็นการหาความน่าจะเป็นของ 2 เหตุการณ์ใด ๆ ที่มีความสัมพันธ์กัน โดยที่เหตุการณ์หนึ่งเกิดขึ้นก่อน แล้วหาความน่าจะเป็นของอีกเหตุการณ์หนึ่ง

สามารถเขียนแทนด้วยสัญลักษณ์

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

Independence

ถ้า X และ Y เป็นอิสระต่อกันจะสามารถเขียนอยู่ใน 2 รูปแบบดังต่อไปนี้

1. $P(X \cap Y) = P(X,Y) = P(X)P(Y)$ ซึ่งเรียกว่า Joint probability โดยจะเป็นการคูณกันของ 2 distributions เพื่อจะได้ไม่ต้องไปหาทุก combination ของค่า X และค่า Y
2. $P(X \mid Y) = P(X)$

โดยการเขียนให้อยู่ใน 2 รูปแบบนี้ก็เพื่อลดความซับซ้อนลงและสร้างแบบจำลองได้ง่ายขึ้น

Naïve Bayes Algorithm ที่จะนำมาทำเป็นโมเดลที่จะใช้ในการทำ Classification มีเป้าหมายคือการ train classifier เพื่อให้ได้ค่า output ออกมา โดยใช้กฎของ Bayes ดังต่อไปนี้

$$P(Y = y_k|X_1, \dots, X_n) = \frac{P(Y = y_k)P(X_1, \dots, X_n|Y = y_k)}{\sum_j P(Y = y_j)P(X_1, \dots, X_n|Y = y_j)}$$

ซึ่งสามารถนำ Conditional Independence มาปรับใช้ได้ดังนี้

$$P(Y = y_k|X_1, \dots, X_n) = \frac{P(Y = y_k)\prod_i P(X_i|Y = y_k)}{\sum_j P(Y = y_k)\prod_i P(X_i|Y = y_j)}$$

จากสมการดังกล่าว เป็นการหาความน่าจะเป็นของเหตุการณ์ แต่ในการทำ Classification ไม่จำเป็นต้องหาความน่าจะเป็น เพราะสามารถทำนายจากการหาค่าที่มากที่สุดได้ทันที ด้วยสมการดังต่อไปนี้

$$Y = argmax_{y_k} P(Y = y_k) \prod_i P(X_i|Y = y_k)$$

โดยวิธีการหาค่าที่มากที่สุดของ Naïve Bayes Algorithm จะแบ่งตามประเภทของข้อมูล มี 2 ประเภท ดังนี้

Discrete

สำหรับ discrete เป็นการหาโดยใช้ตัวแปรสุ่มแบบไม่ต่อเนื่อง สมมติให้ตัวแปร X เป็น discrete variable ที่มีค่าที่เป็นไปได้ J ตัวและตัวแปร Y เป็น discrete variable ที่มีค่าที่เป็นไปได้ K ตัว สามารถหาค่าที่มากที่สุดได้จากการประมาณค่าพารามิเตอร์ 2 ตัว คือ

- $\hat{\theta} = \hat{P}(X_i = x_{ij}|Y = y_k)$
- $\pi_K \equiv P(Y = y_k)$

โดยสามารถประมาณค่าพารามิเตอร์ทั้ง 2 ได้จากการทำ Maximum likelihood estimates (MLE)

$$\hat{\theta} = \hat{P}(X_i = x_{ij}|Y = y_k) = \frac{\#D\{X_i = x_{ij} \text{ and } Y = y_k\}}{\#D\{Y = y_k\}}$$

เป็นการดูจำนวนที่เป็น X และ Y ส่วนด้วยจำนวน Y ทั้งหมด

$$\hat{\pi}_k = \hat{P}(Y = y_k) = \frac{\#D\{Y = y_k\}}{|D|}$$

เป็นการดูจำนวน Y ทั้งหมดส่วนด้วยจำนวนของ Output

Continuous

สำหรับ continuous เป็นการหาโดยใช้ตัวแปรสุ่มแบบต่อเนื่อง สมมติให้ตัวแปร X เป็น continuous variable และตัวแปร Y เป็น discrete variable จะสามารถหาการกระจายตัวของ X ได้จาก Gaussian Distribution โดย Gaussian หาได้จากค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานจากตัวแปร X และ Y มีสมการดังนี้

$$\mu_{ik} = E[X_i|Y = y_k]$$
$$\sigma_{ik}^2 = E[((X_i - \mu_{ik})^2 \mid Y = y_k)]$$

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