Naive Bayes

Naive Baye's Algorithm (classification)

Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem. Despite its simplicity, Naive Bayes is surprisingly effective for a wide range of classification tasks, especially when dealing with large feature spaces.

Bayes Theorem

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

∴ $P(A/B)$ → Prob. of A, given B has occurred

$P(B/A)$ → Prob. of B, given A has occurred

$P(A)$, $P(B)$ → Independent Event of A and B

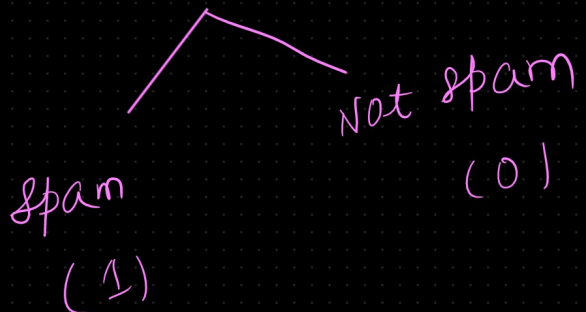
→ Naive Bayes is a straightforward and efficient classification algorithm based on Bayes' theorem and the assumption of feature independence

Naive → Assumption

The "Naive" in Naive Bayes comes from the assumption of feature independence. It assumes that the presence of a particular feature in a class is independent of the presence of other features.

Intution

Dataset → Email classification



→ Aim: To predict whether a new email is spam or Not

Training →

$P(word / spam)$ → Prob. of a word occurring in spam emails

$P(word / not\text{-}spam)$ → " " " in a not spam email

Prediction

Bayes Theorem

$$P(spam/email) = \frac{P(spam) \cdot P(email / spam)}{P(email)}$$

$$\therefore P(email / spam) = P(word_1 / spam) \times P(word_2 / spam)$$

$$P(not\ spam / email) = \frac{P(not\ spam) \cdot P(email / not\ spam)}{P(email)}$$

Decision : $P(spam/email) > P(not\text{-}spam/email)$

$\therefore$ Email → spam

Key Advantages:

- Simple and easy to implement.
- Works well with high-dimensional data.
- Requires relatively few training data.

Key Limitations:

- Strong assumption of feature independence, which may not hold true in all cases.
- May not perform well if features are correlated.