■ NAIVE BAYES — FULL INTERVIEW & REVISION NOTES

■ 1. Bayes Theorem (Foundation of Naive Bayes)

Bayes theorem tells us how to update our belief based on new evidence.

$P(y|x) = P(x|y) * P(y) / P(x)$

Where:

- $P(y) \rightarrow$ Prior: probability of class before seeing data
- $P(x|y) \rightarrow$ Likelihood: probability of features given class
- $P(x) \rightarrow$ Evidence (same for all classes)
- $P(y|x) \rightarrow$ Posterior: updated probability of class after seeing data

Why denominator is ignored during classification?

Because it is same for all classes.

We only compare numerators:

$P(y|x) \propto P(x|y) * P(y)$

■ 2. Naive Bayes Classifier

A probabilistic classifier that applies Bayes rule + Naive assumption.

■ 3. Why called "Naive"?

Because it assumes:

$x_1, x_2, x_3, \ldots, x_n$ are independent given class y.

This assumption is usually false in real life → that's why it's called "Naive".

But surprisingly, it works very well in many real tasks.

■ 4. Naive Bayes Formula

If features are independent:

$$P(y \mid x_1, x_2, \ldots, x_n) \propto P(y) \times \prod P(x_i \mid y)$$

Prediction:

■ = $\text{argmax}_y \left[P(y) \times \prod P(x_i \mid y) \right]$

■ 5. Types of Naive Bayes (Important for Interviews)

1. Gaussian Naive Bayes (for continuous data)

Assumes each feature follows a Gaussian/Normal distribution:

$$P(x|y) = (1 / \sqrt{2\pi\sigma^2}) * \exp(-(x-\mu)^2 / (2\sigma^2))$$

Use when:

- features are continuous
- shape looks bell-curve like

Examples:

- Iris dataset

• Medical numeric features

2. Multinomial Naive Bayes (for counts)

Used when features are counts / frequencies, like:

• word counts

• number of clicks

Used heavily in:

• Text classification

• Spam detection

3. Bernoulli Naive Bayes (for binary features)

Used when features are 0/1 (present/absent).

■ 6. Advantages of Naive Bayes

• Simple, fast, reliable

• Works well with small datasets

• Excellent for high-dimensional data (text)

• Needs very little training data

• Often beats more complex models

• Produces probability output

■ 7. Disadvantages of Naive Bayes

• Assumes independence of features

• Zero-probability problem (fixed by Laplace smoothing)

• Not good for non-Gaussian numeric data

• Fails when features are correlated

■ 8. Laplace Smoothing (VERY Important)

Fixes the zero-probability issue:

$P(x_i \mid y) = (\text{count} + 1) / (\text{total} + V)$

Where V = vocabulary size.

■ 9. When to Use Naive Bayes

• Text classification

• Spam filtering

• Sentiment analysis

• Medical diagnosis

• Real-time predictions

■ 10. When NOT to Use Naive Bayes

• Features highly correlated

• Data not independent

• Numeric data not Gaussian

• Very high accuracy required on continuous values

■ 11. End-to-End Flow of Naive Bayes

1. Calculate class priors P(y)

2. Compute P(xi | y) for each feature & class

3. Multiply them: P(y) × Π P(xi | y)

4. Ignore denominator (same)

5. Choose class with highest value

6. (Optional) Normalize to get actual probabilities

■ 12. One-Line Summary for Interviews

Naive Bayes = Bayes theorem + feature independence + (Gaussian, Bernoulli, Multinomial models).

■ 13. Rapid-Fire Revision

- Bayes rule
- Ignore denominator
- Multiply likelihoods
- Gaussian/Bernoulli/Multinomial
- Laplace smoothing
- Great for text classification