Prior Probabilities

Prior probabilities in Naive Bayes represent how likely each class is before seeing any features they can be (1) empirical (based on class frequencies), (2) uniform (all classes equally likely), or (3) custom

21 OCT, 2025

BAYESIANBRAT*

Step-by-Step Understanding of Priors in Naive Bayes

We will go step by step with actual numbers:

1. what a **prior** is,
2. how we **use it** in Naive Bayes (multiply by likelihood), and
3. three concrete numeric scenarios (empirical priors, uniform priors, and a heavily biased prior).

1) What is a Prior Probability?

- **Prior** $P(Y = y)$ is your belief about how common each class y is *before* seeing the features of the current instance.
- Common choices:
  - **Empirical prior**: use class frequency in training data,
  $$P(Y = y) = \frac{N(Y = y)}{N_{\text{total}}}.$$
  - **Uniform prior**: assume all classes equally likely.
  - **Custom prior**: set by domain knowledge or to reflect costs/importance.

2) How Prior is Used in Naive Bayes

For a given instance (e.g. `Weather = wet`, `Surface = medium`), Naive Bayes computes for each class y an unnormalized score:

$$\text{score}(y) = P(\text{features} \mid Y = y) \times P(Y = y).$$

Then normalize:

$$P(Y = y \mid \text{features}) = \frac{\text{score}(y)}{\sum_{y'} \text{score}(y')}.$$

Thus, the prior multiplies the likelihood and scales the class score.

3) Concrete Numbers (from the smoothed example)

Smoothed likelihoods:

$$\begin{aligned} P(\text{wet}, \text{med} \mid \text{major}) &= 0.375 \times 0.625 = 0.234375, \\ P(\text{wet}, \text{med} \mid \text{minor}) &= 0.0625 \times 0.5625 = 0.03515625, \\ P(\text{wet}, \text{med} \mid \text{none}) &= 0.2 \times 0.4 = 0.08. \end{aligned}$$

We now plug in different priors.

A — Empirical Priors

From the dataset: Major=5, Minor=13, None=2 (total 20):

$$P_{\text{emp}}(\text{maj}) = \tfrac{5}{20} = 0.25, \quad P_{\text{emp}}(\text{min}) = \tfrac{13}{20} = 0.65, \quad P_{\text{emp}}(\text{none}) = \tfrac{2}{20} = 0.10.$$

Scores:

$$\begin{aligned} \text{score(maj)} &= 0.234375 \times 0.25 = 0.05859375, \\ \text{score(min)} &= 0.03515625 \times 0.65 = 0.0228515625, \\ \text{score(none)} &= 0.08 \times 0.10 = 0.008. \end{aligned}$$

Sum of scores = 0.0894453125.
Normalized posteriors:

$$\begin{aligned} P(\text{maj} \mid \text{features}) &\approx 0.6551, \\ P(\text{min} \mid \text{features}) &\approx 0.2555, \\ P(\text{none} \mid \text{features}) &\approx 0.0894. \end{aligned}$$

Interpretation: Empirical priors (favoring `minor`) shift the posteriors, but `major` still wins due to stronger likelihood.

B — Uniform Priors

Set each prior = 1/3.
Scores:

$$\begin{aligned} \text{score(maj)} &= 0.234375 \times \tfrac{1}{3} = 0.078125, \\ \text{score(min)} &= 0.03515625 \times \tfrac{1}{3} = 0.01171875, \\ \text{score(none)} &= 0.08 \times \tfrac{1}{3} \approx 0.02667. \end{aligned}$$

Sum $\approx$ 0.11651.
Normalized posteriors:

$$\begin{aligned} P(\text{maj} \mid \text{features}) &\approx 0.6705, \\ P(\text{min} \mid \text{features}) &\approx 0.1006, \\ P(\text{none} \mid \text{features}) &\approx 0.2289. \end{aligned}$$

Interpretation: With uniform priors, `major` gains more probability, since `minor`'s empirical advantage is removed.

C — Strongly Biased Prior

Suppose: Major=0.05, Minor=0.90, None=0.05.
Scores:

$$\begin{aligned}
\text{score(maj)} &= 0.234375 \times 0.05 = 0.01171875, \\
\text{score(min)} &= 0.03515625 \times 0.90 = 0.031640625, \\
\text{score(none)} &= 0.08 \times 0.05 = 0.004.
\end{aligned}$$

Sum = 0.047359375.
Normalized posteriors:

$$\begin{aligned}
P(\text{maj} \mid \text{features}) &\approx 0.2474, \\
P(\text{min} \mid \text{features}) &\approx 0.6681, \\
P(\text{none} \mid \text{features}) &\approx 0.0845.
\end{aligned}$$

Interpretation: Strong prior on `minor` flips the decision, despite weaker likelihood.

4) Intuition Summary

- Prior acts like a *multiplier.*
- If two classes have similar likelihoods, the prior decides.
- Strong likelihood can overcome modest prior bias.
- Strong priors can dominate and override evidence.

5) Rule of Thumb

- Use empirical priors (class fractions) if training data reflects reality.
- Use uniform priors if you want neutrality.
- Use custom priors if you have domain knowledge or want to adjust costs/imbalance.

6) Odds Perspective

Posterior odds between two classes:

$$\frac{P(\text{maj} \mid x)}{P(\text{min} \mid x)} = \frac{P(\text{maj})}{P(\text{min})} \times \frac{P(x \mid \text{maj})}{P(x \mid \text{min})}.$$

Interpretation: "Prior odds" $\times$ "likelihood ratio" $=$ "posterior odds".