Naive Bayes Nuances

Parth Bansal

June 16, 2025

Maximum Likelihood Estimation for the Bernoulli Naive Bayes Model

We have training data

$$\{(x^{(i)}, y^{(i)})\}_{i=1}^{n},$$

where each $y^{(i)} \in \{0, 1\}$ and each feature vector

$$x^{(i)} = (x_1^{(i)}, \dots, x_d^{(i)})$$

with $x_j^{(i)} \in \{0, 1\}$.

Our model parameters are:

$$\begin{aligned}
&\phi_y = p(y = 1), \quad p(y = 0) = 1 - \phi_y, \\
&\phi_{j|1} = p(x_j = 1 \mid y = 1), \quad p(x_j = 0 \mid y = 1) = 1 - \phi_{j|1}, \\
&\phi_{j|0} = p(x_j = 1 \mid y = 0), \quad p(x_j = 0 \mid y = 0) = 1 - \phi_{j|0}.
\end{aligned}$$

By independence of the x_j given y, the joint for a single example is

$$p\big(x^{(i)}, y^{(i)}\big) = [\phi_y]^{y^{(i)}} [1 - \phi_y]^{1 - y^{(i)}} \prod_{j=1}^{d} [\phi_{j|y^{(i)}}]^{x_j^{(i)}} [1 - \phi_{j|y^{(i)}}]^{1 - x_j^{(i)}}.$$

Hence the full data likelihood is

$$\mathcal{L}(\phi_y, \{\phi_{j|1}, \phi_{j|0}\}) = \prod_{i=1}^{n} p\big(x^{(i)}, y^{(i)}\big).$$

Log-Likelihood

Define the log-likelihood:

$$\ell = \log \mathcal{L} = \sum_{i=1}^{n} \big[y^{(i)} \log \phi_y + (1 - y^{(i)}) \log(1 - \phi_y) \big] + \sum_{j=1}^{d} \sum_{i=1}^{n} \big[x_j^{(i)} \log \phi_{j|y^{(i)}} + (1 - x_j^{(i)}) \log(1 - \phi_{j|y^{(i)}}) \big].$$

Parameters separate, so we maximize each group.

MLE for ϕ_y

Extract terms in ℓ involving ϕ_y:

$$\ell(\phi_y) = \sum_{i=1}^{n} \big[y^{(i)} \log \phi_y + (1 - y^{(i)}) \log(1 - \phi_y) \big].$$

Differentiate and set to zero:

$$\begin{aligned} \frac{\partial \ell}{\partial \phi_y} &= \sum_{i=1}^{n} \Big[\frac{y^{(i)}}{\phi_y} - \frac{1 - y^{(i)}}{1 - \phi_y} \Big] = 0 \\ &\implies \sum_{i=1}^{n} y^{(i)}(1 - \phi_y) - \sum_{i=1}^{n} (1 - y^{(i)}) \phi_y = 0 \implies \phi_y = \frac{1}{n} \sum_{i=1}^{n} y^{(i)}. \end{aligned}$$

MLE for $\phi_{j|1}$

Terms involving $\phi_{j|1}$ (only $y^{(i)} = 1$):

$$\ell(\phi_{j|1}) = \sum_{i:y^{(i)}=1} [x_j^{(i)} \log \phi_{j|1} + (1 - x_j^{(i)}) \log(1 - \phi_{j|1})].$$

Differentiate:

$$\begin{aligned} \frac{\partial \ell}{\partial \phi_{j|1}} &= \sum_{i:y^{(i)}=1} \Big[\frac{x_j^{(i)}}{\phi_{j|1}} - \frac{1 - x_j^{(i)}}{1 - \phi_{j|1}} \Big] = 0 \\ &\implies \sum_{i:y^{(i)}=1} x_j^{(i)} - \big(\sum_{i:y^{(i)}=1} 1 \big) \phi_{j|1} = 0 \implies \phi_{j|1} = \frac{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 1\} x_j^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 1\}}. \end{aligned}$$

MLE for $\phi_{j|0}$

Similarly for $y^{(i)} = 0$:

$$\phi_{j|0} = \frac{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 0\} x_j^{(i)}}{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 0\}}.$$

Summary

$$\begin{aligned} \hat{\phi}_y &= \frac{1}{n} \sum_{i=1}^{n} y^{(i)}, \\ \hat{\phi}_{j|1} &= \frac{\sum_{i=1}^{n} \mathbf{1}\{x_j^{(i)} = 1 \wedge y^{(i)} = 1\}}{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 1\}}, \\ \hat{\phi}_{j|0} &= \frac{\sum_{i=1}^{n} \mathbf{1}\{x_j^{(i)} = 1 \wedge y^{(i)} = 0\}}{\sum_{i=1}^{n} \mathbf{1}\{y^{(i)} = 0\}}. \end{aligned}$$