Binary Naive Bayes in Exponential Family Form

Yaroslav Bulatov

April 29, 2006

Consider Naive Bayes classifier over binary domain. We want to predict $Y \in \{0,1\}$ given $X_{1,\dots,n} \in \{0,1\}^n$.

We have

$$P(Y = y|X_{1,\dots,n} = x_{1,\dots,n}) = P(Y = y)\prod_{i=1}^{n} P(X_i = x_i) \tag{1}$$

For compactness introduce the following notation:

$$p_0 = P(Y = 1)$$
$$p_{i0} = P(X_i = 1|Y = 0)$$
$$p_{i1} = P(X_i = 1|Y = 1)$$

Then we can write our probabilities as follows

$$P(X_i = x_i|Y = y) = p_{i0}^{(1-y)x_i}(1-p_{i0})^{(1-y)(1-x_i)}p_{i1}^{yx_i(1-p_{i1})}(1-p_{i1})^{y(1-x_i)} \tag{2}$$

Similary

$$P(Y = y) = p_0^y(1-p_0)^{1-y} \tag{3}$$

Substituting into original formula we get

$$P(Y = y, X_{1,\dots,n}) = p_0^y(1-p_0)^{1-y}\prod_{i=1}^{n} p_{i0}^{(1-y)x_i}(1-p_{i0})^{(1-y)(1-x_i)}p_{i1}^{yx_i}(1-p_{i1})^{y(1-x_i)} \tag{4}$$

It's easier to work with logs, so

$\log P(Y = y, X_{1,\ldots,n} = x_{1,\ldots,n}) =$
$= y \log p_0 + (1-y)\log(1-p_0) +$
$+ \sum_{i=1}^{n} (1-y) x_i \log p_{i0} + (1-y)(1-x_i)\log(1-p_{i0}) +$
$+ \sum_{i=1}^{n} y x_i \log p_{i1} + y(1-x_i)\log(1-p_{i1})$
Expand and group like terms to get

$$\begin{aligned}
\log P(Y = y, X_{1,\ldots,n} = x_{1,\ldots,n}) &= \log(1-p_0) + \sum_{i=1}^{n} \log(1-p_i0) && (5)\\
&+ y\left[\log \frac{p_0}{1-p_0} + \sum_{i=1}^{n} \log \frac{1-p_{i1}}{1-p_{i0}}\right] && (6)\\
&+ \sum_{i=1}^{n} x_i \log \frac{p_{i0}}{1-p_{i0}} && (7)\\
&+ \sum_{i=1}^{n} x_i y \log \frac{1-p_{i0}}{p_{i0}} \frac{p_{i1}}{1-p_{i1}} && (8)
\end{aligned}$$

Changing parameters we write

$$P(Y = y, X_{1,\ldots,n} = x_{1,\ldots,n}) = \exp(\theta_0 y + \sum_{i=1}^{n} \phi_i x_i + \sum_{i=1}^{n} \theta_i x_i y - A(p)) \tag{9}$$

Where

$$\begin{aligned}
\theta_0 &= \log \frac{p_0}{1-p_0} + \sum_{i=1}^{n} \log \frac{1-p_{i1}}{1-p_{i0}} && (10)\\
\phi_i &= \log \frac{p_{i0}}{1-p_{i0}}, && (11)\\
\theta_i &= \log \frac{1-p_{i0}}{p_{i0}} \frac{p_{i1}}{1-p_{i1}} && (12)\\
A(p) &= -\log(1-p_0) - \sum_{i=1}^{n} \log(1-p_{i0}) && (13)
\end{aligned}$$

This isn't quite the exponential family form because the normalizer, $A(p)$, needs to be in terms of our new parameters. First we solve for p_i's to get

$$p_0 = \frac{e^{\theta_0} \prod_{i=1}^{n} \frac{1+e^{\phi_i+\theta_i}}{1+e^{\phi_i}}}{1 + e^{\theta_0} \prod_{i=1}^{n} \frac{1_e^{\phi_i+\theta_i}}{1+e^{\phi_i}}} \tag{14}$$

$$
\begin{aligned}
p_{i0} &= \frac{e^{\phi_i}}{1+e^{\phi_i}} && (15)\\
p_{i1} &= \frac{e^{\phi_i+\theta_i}}{1+e^{\phi_i+\theta_i}} && (16)
\end{aligned}
$$

Substituting into $A(p)$ and simplifying get

$$
A = \log\left(\prod_{i=1}^{n}(1+e^{\phi_i}) + e^{\theta_0}\prod(1+e^{\phi_i+\theta_i})\right) \tag{17}
$$

Finally our equation becomes

$$
P(Y=y, X_{1,\ldots,n}=x_{1,\ldots,n}) = \frac{\exp(\theta_0 y + \sum_{i=1}^{n}\phi_i x_i + \sum_{i=1}^{n}\theta_i x_i y)}{\prod_{i=1}^{n}(1+e^{\phi_i}) + e^{\theta_0}\prod(1+e^{\phi_i+\theta_i})} \tag{18}
$$

It can be useful to rewrite this as a product of conditional and marginal distributions:

$$
P(Y, X_{1,\ldots,n}) = \underbrace{\frac{\exp(\theta_0 + \sum_{i=1}^{n}\theta_i x_i)^y}{1+\exp(\theta_0+\sum_{i=1}^{n}\theta_i x_i)}}_{\text{conditional}} \cdot \underbrace{\frac{\exp(\sum_{i=1}^{n}\phi_i x_i) + \exp(\sum_{i=1}^{n}(\phi_i+\theta_i)x_i + \theta_0)}{\prod_{i=1}^{n}(1+e^{\phi_i}) + e^{\theta_0}\prod(1+e^{\phi_i+\theta_i})}}_{\text{marginal}} \tag{19}
$$

You can see that the conditional part is equivalent to logistic regression. Another thing to note is that since conditional part involves only θ's, for every logistic regression model, there's a corresponding Naive Bayes model (just set ϕ's to be anything).