# NAÏVE BAYES

$p$ features, categorical response $y \in C = \{C_1, \dots, C_K\}$

$$(x_1, y_1), \dots, (x_n, y_n), \quad x_i = (x_{i1}, \dots, x_{ip})$$

Bayes rule:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \propto p(x|y)p(y)$$

$$\widehat{p}(y) := \frac{\#[y_i = y]}{n}$$

$$\widehat{p}(x|y) := \frac{\#[(x_i, y_i) = (x, y)]}{\#[(x_i, y_i) = (*, y)]}$$

$$\widehat{p}(y|x) := \frac{\widehat{p}(x|y)\widehat{p}(y)}{\widehat{p}(x)} \propto \widehat{p}(x|y)\widehat{p}(y)$$

Decision rule: Assign $x_0$ to class $i$ if

$$\widehat{p}(y|x_0) > \widehat{p}(y'|x_0) \quad \text{for all} \quad y' \neq y.$$

Let $\xi \in \mathbb{R}^p$, $c \in C$

$$\widehat{p}(\xi|c) = \frac{\#[(x_i, y_i) = (\xi, c)]}{\#[(x_i, y_i) = (*, c)]} = \frac{\#[(x_{i1}, \dots, x_{ip}, y_i) = (\xi_1, \dots, \xi_p, c)]}{\#[(x_i, y_i) = (*, c)]}$$

Suppose $X_1, \dots, X_p$ are conditionally independent given $Y$. Then

$$\widehat{p}(\xi|c) = p(\xi_1|c) \cdots p(\xi_p|c)$$

$$\widehat{p}(\xi_j|c) := \frac{\#[(x_{ij}, y_i) = (\xi, c)]}{\#[(x_{ij}, y_i) = (*, c)]}$$

## 1. Take two

Let $\boldsymbol{X} = (X_1, \dots, X_p)$ be a random vector and let $Y$ be a categorical random variable taking values in $\{1, \dots, K\}$. Given a realization $\boldsymbol{x}$ of $X$, define

$$f(\boldsymbol{x}) = \operatorname*{argmax}_y p(y|\boldsymbol{x}).$$

By Bayes' rule,

$$f(\boldsymbol{x}) = \operatorname*{argmax}_y p(\boldsymbol{x}|y)p(y).$$

Assume that $Y$ has a categorical distribution with mass function

$$p(y) = \pi_1^{[y=1]} \cdots \pi_K^{[y=K]},$$

where $\pi_1 + \cdots \pi_K = 1$ and the symbolk $[y = k]$ stands for 1 is $y = k$ and for 0 if $y \neq k$. Equivalently,

$$p(Y = k) = \pi_k, \qquad 1 \leq k \leq K.$$

We estimate $p(\boldsymbol{x}|y)$ and $p(y)$ using training data, $(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(n)}, y^{(n)})$.

$$p_{jy} = p(X_j = 1|Y = y)$$

Then $\widehat{p}_{yj}$, where

$$p_{jy} \approx \widehat{p}_{jy} := \frac{\#\{i : x_j^{(i)} = 1 \text{ and } y^{(i)} = y\}}{\#\{i : y^{(i)} = y\}}.$$

Note that

$$\begin{aligned} p(X_j = 0|Y = y) &= 1 - p_{yj} \\ &\approx 1 - \widehat{p}_{yj} \end{aligned}$$

Therefore,

$$\begin{aligned} \widehat{p}(x_j|y) &:= \widehat{p}_{jy}^{x_{yj}} (1 - \widehat{p}_{jy})^{1-x_j} \\ &\approx p_{jy}^{x_{yj}} (1 - p_{jy})^{1-x_j} \\ &= p(X = x_j|Y = y) \end{aligned}$$

Taking logs,

$$\log \widehat{p}(x_j|y) = x_j \log \widehat{p}_{jy} + (1 - x_j) \log(1 - \widehat{p}_{jy})$$

$$\sum_{j=1}^{p} \log p(x_j|y) =$$

$$I(\boldsymbol{y} = y) \cdot \boldsymbol{x}_j = \#\{i : x_j^{(i)} = 1 \text{ and } y^{(i)} = y\}$$
$$I(\boldsymbol{y} = y) \cdot I(\boldsymbol{y} = y) = \#\{i : y^{(i)} = y\}$$

$$\widehat{p}_{jy} = \frac{I(\boldsymbol{y} = y) \cdot \boldsymbol{x}_j}{I(\boldsymbol{y} = y) \cdot I(\boldsymbol{y} = y)}$$

$$\widehat{p}_y = \frac{I(\boldsymbol{y} = y) \cdot \boldsymbol{x}}{I(\boldsymbol{y} = y) \cdot I(\boldsymbol{y} = y)}$$