

MLE of Naive Bayes

Data: $y \in \{0,1\} \in \{H,T\}$ where real = 1 and fake = 0
 $\vec{x} \in \{0,1\}^M$ where $M = \#$ of words in vocabulary

① Model: $y \sim \text{Bernoulli}(\phi) = p(y|\phi)$
 $x_1 \sim \text{Bernoulli}(\theta_{y,1}) = p(x_1|y, \theta)$
 $x_2 \sim \text{Bernoulli}(\theta_{y,2}) = p(x_2|y, \theta)$
 \vdots
 $x_M \sim \text{Bernoulli}(\theta_{y,M}) = p(x_M|y, \theta)$

$\phi \in [0,1]$
 $\theta = \begin{bmatrix} \theta_{H,1} & \theta_{H,2} & \dots & \theta_{H,M} \\ \theta_{T,1} & \theta_{T,2} & \dots & \theta_{T,M} \end{bmatrix}$ $\in [0,1]$

Joint:
 $p(\vec{x}, y) = p(y, x_1, x_2, \dots, x_M) = p(y|\phi) p(x_1|y, \theta) \dots p(x_M|y, \theta)$
 $= p(y|\phi) \prod_{m=1}^M p(x_m|y, \theta_{H,m}, \theta_{T,m})$
 $= \underbrace{\phi^y (1-\phi)^{(1-y)}}_{p(y)} \underbrace{\prod_{m=1}^M \theta_{y,m}^{x_m} (1-\theta_{y,m})^{(1-x_m)}}_{p(\vec{x}|y)}$

Def: two r.v.s X, Y are conditionally independent given r.v. Z , written $X \perp\!\!\!\perp Y | Z$, iff $P(X, Y | Z) = P(X | Z) P(Y | Z)$

Naive Bayes Assumption
 $p(\vec{x}|y) = \prod_{m=1}^M p(x_m|y)$
 $\Rightarrow x_q$ and x_r are cond. indep. given y

② log-likelihood

$l(\phi, \theta) = \log \prod_{i=1}^N p(x^{(i)}, y^{(i)} | \phi, \theta)$
 $= \sum_{i=1}^N \log p(y^{(i)} | \phi) + \sum_{m=1}^M \log p(x_m^{(i)} | y^{(i)}, \theta)$
 $= N_{y=H} \log(\phi) + N_{y=T} \log(1-\phi) +$

$N_{y=H} = (\# y^{(i)} = H)$
 $N_{y=T} = (\# y^{(i)} = T)$

$N_{x_m=1, y=H} = \#$ times that $x_m^{(i)} = 1$ and $y^{(i)} = H$

$\sum_{m=1}^M N_{x_m=1, y=H} \log(\theta_{H,m}) + N_{x_m=0, y=H} \log(1-\theta_{H,m})$ $y^{(i)} = H$
 $\sum_{m=1}^M N_{x_m=1, y=T} \log(\theta_{T,m}) + N_{x_m=0, y=T} \log(1-\theta_{T,m})$ $y^{(i)} = T$
 $x_m^{(i)} = 1$ $x_m^{(i)} = 0$

$$X_M^{(1)} = 1$$

$$X_M^{(2)} = 0$$

Case a) ϕ b) Θ elements

③ a) Compute partial derivative wrt ϕ

$$\frac{\partial \ell(\phi, \theta)}{\partial \phi} = \frac{N_{y=H}}{\phi} - \frac{N_{y=T}}{1-\phi}$$

big gold coin



④ a) Set to zero and solve for ϕ

$$\phi_{MLE} = \frac{N_{y=H}}{N_{y=H} + N_{y=T}} = \frac{N_{y=H}}{N}$$

③ b) Compute partial derivative wrt $\Theta_{H,j}$

$$\frac{\partial \ell(\phi, \theta)}{\partial \Theta_{H,j}} = \frac{N_{X_j=1, y=H}}{\Theta_{H,j}} - \frac{N_{X_j=0, y=H}}{1-\Theta_{H,j}}$$

red coins



or

blue coins



④ b) Set to zero and solve for $\Theta_{H,j}$

$$\Theta_{H,j}^{MLE} = \frac{N_{X_j=1, y=H}}{N_{X_j=1, y=H} + N_{X_j=0, y=H}} = \frac{N_{X_j=1, y=H}}{N_{y=H}}$$

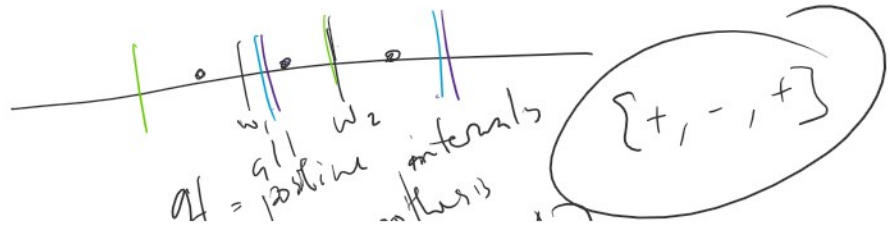
OH

$H[S]$ = set of all labelings H can give to $S = \{$

$|H[S]|$ = # labelings that H can give to S

- $[- + -]$
- $[+ - -]$
- $[- + +]$
- $[+ + +]$
- $[- - -]$
- $[- - +]$

$2^{|S|}$ = # of possible labelings of S



$H =$ all positive integer
hypothesis
function in \mathbb{Z}

$H_1 =$ all depth 3 DTs
 $H_2 =$ all depth 4 DTs

