Section B

Monday, March 20, 2023 12:29 PM

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:**

$y \sim \text{Bernoulli}(\phi) = p(y|\phi)$ $\quad \phi \in [0,1]$

$x_1 \sim \text{Bernoulli}(\theta_{y,1}) = p(x_1|y,\theta)$

$x_2 \sim \text{Bernoulli}(\theta_{y,2}) = p(x_2|y,\theta)$

$\vdots$

$x_M \sim \text{Bernoulli}(\theta_{y,M}) = p(x_M|y,\theta)$

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

Joint:

$$p(\vec{x},y) = p(y,x_1,x_2,\dots,x_M) = p(y|\phi)\,p(x_1|y,\theta)\cdots p(x_M|y,\theta)$$

$$= p(y|\phi)\prod_{m=1}^{M} p(x_m|y,\theta_{H,m},\theta_{T,m})$$

$$= \underbrace{\phi^{y}(1-\phi)^{(1-y)}}_{p(y)} \underbrace{\prod_{m=1}^{M} \theta_{y,m}^{(x_m)}(1-\theta_{y,m})^{(1-x_m)}}_{p(\vec{x}|y)}$$

Def: two r.v.s X, Y are conditionally independent given r.v. Z, written $X \perp\!\!\!\perp Y \mid Z$, 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**

$$\ell(\phi,\theta) = \log \prod_{i=1}^{N} p(x^{(i)},y^{(i)}|\phi,\theta)$$

$$= \sum_{i=1}^{N} \boxed{\log p(y^{(i)}|\phi)} + \sum_{m=1}^{M} \boxed{\log p(x_m^{(i)}|y^{(i)},\theta)}$$

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

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

$$= \boxed{N_{y=H}\log(\phi) + N_{y=T}\log(1-\phi)} +$$

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

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

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

$x_m^{(i)} = 1$ $x_m^{(i)} = 0$

Case a) ϕ b) Θ elements

(3) a) Compute partial derivative wrt ϕ

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

(4) a) Set to zero and solve for ϕ

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

big gold coin

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

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

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

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

red coins

or

blue coins

OH

$\mathcal{H}[S]$ = set of all labelings $\mathcal{H}$ can give to $S = \{ [- + -], [+ - -], [- + +], [+ + +], [- - -], [- - +] \}$

$|\mathcal{H}[S]|$ = # labelings that $\mathcal{H}$ can give to S

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

$\mathcal{H}$ = all positive intervals hypothesis

$[+, -, +]$

all
$\mathcal{H}$ = positive interval
hypothesis
function in 1D

$\mathcal{H}_1$ = all depth 3 DTs
$\mathcal{H}_2$ = all depth 4 DTs

$|\mathcal{H}_2| > |\mathcal{H}_1|$

$VC(\mathcal{H}_2) \geq VC(\mathcal{H}_1)$