ML Assignment - 2

2. Multinomial Naive Bayes.

Classes = {r, u}

Unique words = {Moscow, Russia, Washington, U.S, St, Petersburg, Paul, Syria}

Total Docs = 5

$P(r) = 2/5$

$P(u) = 3/5$

$P(\text{Moscow}|r) = 1+1/8+5 = 2/13$

$P(\text{Russia}|r) = 2+1/8+5 = 3/13$

$P(\text{Washington}|r) = 0+1/8+5 = 1/13$

$P(\text{U.S.}|r) = 0+1/8+5 = 1/13$

$P(\text{st}|r) = 1+1/8+5 = 2/13$

$P(\text{Petersburg}|r) = 1+1/8+5 = 2/13$

$P(\text{Paul}|r) = 0+1/8+5 = 1/13$

$P(\text{Syria}|r) = 0+1/8+5 = 1/13$

$P(\text{Russia}|u) = 1+1/8+9 = 2/17$

$P(\text{Moscow}|u) = 1+1/8+9 = 2/17$

$P(\text{Washington}|u) = 1+1/8+9 = 2/17$

$P(\text{u.s}|u) = 3+1/8+9 = 4/17$

$P(\text{st}|u) = 1+1/8+9 = 2/17$

$P(\text{Petersburg}|u) = 0+1/8+9 = 1/17$

$P(\text{Paul}|u) = 1+1/8+9 = 2/17$

$P(\text{Syria}|u) = 1+1/8+9 = 2/17$

(1) $P(r \mid Moscow, Moscow, st, Russia)$

$= P(r) \cdot P(Moscow \mid r) \cdot P(Moscow \mid r) \cdot P(st \mid r) \cdot P(Russia \mid r)$

$= \frac{2}{5} \cdot \frac{2}{13} \cdot \frac{2}{13} \cdot \frac{2}{13} \cdot \frac{3}{13}$

$= \frac{48}{5 \times 13^2 \times 13^2}$

$= 48/5 \times 169 \times 169 = 48/142805$

$= 0.0003361$

$P(u \mid Moscow, Moscow, st, Russia)$

$= P(u) \cdot P(Moscow \mid u) \cdot P(Moscow \mid u) \cdot P(st \mid u) \cdot P(Russia \mid u)$

$= \frac{3}{5} \cdot \frac{2}{17} \cdot \frac{2}{17} \cdot \frac{2}{17} \cdot \frac{2}{17}$

$= \frac{48}{5 \times 17^2 \times 17^2} = 48/417605 = 0.0001149$

(2) $P(r \mid U.S, St, Petersburg)$

$= P(r) \cdot P(U.S. \mid r) \cdot P(s.t \mid r) \cdot P(Petersburg \mid r)$

$= \frac{2}{5} \cdot \frac{1}{13} \cdot \frac{2}{13} \cdot \frac{2}{13}$

$= 8/65 \cdot 169$

$= 0.00072826$

Scanned by CamScanner

$P(u \mid U.S., St., Petersburg)$

$= P(u) \cdot P(U.S \mid u) \cdot P(St \mid u) \cdot P(Petersburg \mid u)$

$= \frac{3}{5} \cdot \frac{4}{17} \cdot \frac{2}{17} \cdot \frac{1}{17}$

$= \frac{24}{85 \cdot 17^2}$

$= \frac{24}{24565}$

$= 0.00097699$

Doc	Words	Class
1	Moscow, Moscow, St, Russia	r
2	U.S St Petersburg	u

Q. 1 Bayesian Statistics.
($x=0$ for tails $x=1$ for heads)

$$P(x) = \begin{cases} 1-\Theta_1 & x=0 \\ \Theta_1 & x=1 \end{cases}$$

Given table.

	$Y=0$	$Y=1$
$x=0$	Θ_2	$1-\Theta_2$
$x=1$	$1-\Theta_2$	Θ_1

(a) Joint Probability distribution table is given as follows

$$P(x,y) = P(y|x)\,P(x)$$

$$\begin{bmatrix} P(0,0) & P(0,1) \\ P(1,0) & P(1,1) \end{bmatrix} = \begin{array}{c|cc} & y=0 & y=1 \\ \hline x=0 & \Theta_2(1-\Theta_1) & (1-\Theta_2)(1-\Theta_1) \\ x=1 & (1-\Theta_2)\Theta_1 & \Theta_2\Theta_1 \end{array}$$

(b) Given dataset
$x=(1,1,0,1,1,0,0)$ $y=(1,0,0,0,1,0,1)$
Find MLE for Θ_1 & Θ_2.

(P.T.O)

Ans) The MLE of samples (x_i, y_i) given $i = 1$ to n is given as

$$= \sum_i \log P(x_i, y_i)$$

$$= \sum_i \log\left(P(y|x) \cdot P(x)\right)$$

$$= \sum_i \log\left(P(y|x)\right) + \sum_i \log\left(P(x)\right)$$

$$= J_2(\theta_2) + J_1(\theta_1)$$

To minimize $J_1(\theta_1)$

$$J(\theta_1) = \sum_i \log P(x)$$

$$J(\theta_1) = N \log \theta_1 + (n - N) \log(1 - \theta_1)$$

$$N = \sum(x = 1)$$

Differentiating w.r.t θ

$$\frac{dJ}{d\theta} = \frac{N}{\theta_1} - \frac{(n-N)}{1-\theta_1}$$

equating to 0 we obtain

$$\hat{\theta}_1 = \frac{N}{n} \qquad (N = (x=1))$$

$$J_2(\theta_2) = \sum \log P(y_i | x_i)$$

$$= N \log \theta_2 + (n - N) \log(1-\theta_2)$$

where $N =$ number of $(x = y)$

Differentiating w.r.t to θ_2

$$\frac{dJ}{d\theta_2} = \frac{N}{\theta_2} - \frac{(n-N)}{1-\theta_2}$$

equating to 0 we get

$$\hat{\theta}_2 = \frac{N}{n} \quad \text{wher } (N = (x = y))$$

For the given data set.

$$\hat{\theta}_1 = \frac{4}{7}$$

as there 4 coin toses where $x = 1$

$$\hat{\theta}_2 = 4/7$$

as there are 4 instances when y is same as x.

Scanned by CamScanner

Calculate $P(D \mid \hat{\Theta}, M_2)$

$$D = \{(x_i, y_i) \quad i = 1 \dots n\}$$

$$P(D \mid \hat{\theta}, M_2) = \prod_i P(y_i \mid x_i) \cdot P(x_i)$$

$$\begin{aligned} P(x, y \mid \hat{\Theta}, M_2) &= P(1,1 \mid \hat{\Theta}) \\ &\times P(1,0 \mid \hat{\Theta}) \\ &\times P(0,0 \mid \hat{\Theta}) \\ &\times P(1,0 \mid \hat{\Theta}) \\ &\times P(1,1 \mid \hat{\Theta}) \\ &\times P(0,0 \mid \hat{\Theta}) \\ &\times P(0,1 \mid \hat{\Theta}) \end{aligned}$$

$$\begin{aligned} &= (\hat{\Theta}_2 \hat{\Theta}_1) \times ((1 - \hat{\Theta}_2)\hat{\Theta}_1) \times (\hat{\Theta}_2 (1 - \hat{\Theta}_1)) \\ &\times ((1 - \hat{\Theta}_2)(\hat{\Theta}_1)) \times (\hat{\Theta}_2 \hat{\Theta}_1) \times (\hat{\Theta}_2 (1 - \hat{\Theta}_1)) \\ &\times ((1 - \hat{\Theta}_2)(1 - \hat{\Theta}_1)) \end{aligned}$$

$$\begin{aligned} &= 4/7 \times 4/7 \times 3/7 \times 4/7 \times 4/7 \times 3/7 \times 3/7 \\ &\times 4/7 \times 4/7 \times 4/7 \times 4/7 \times 3/7 \times 3/7 \times 3/7 \end{aligned}$$

$$= (4/8)^8 \times (3/7)^6.$$