Assignment 4

Information Retrieval and Text Mining 20/21

Team Members:

1. Suhas Devendrakeerti Sangolli: (M. No: 3437641)
2. Tushar Rajendra Balihalli: (M. No: 3437638)
3. Stefan Truong: (M. No: 3338287)

Assignment - 04

Task-1). doc = "happy new year celebrations". [∵ Given]

Now, $P(c1|doc) \propto P(c1) \cdot \prod_{w \in doc} p(w|c1)$

$\Rightarrow P(c1|doc) \propto \left[\frac{3}{5} \cdot p(happy|c1) \cdot p(new|c1) \cdot p(year|c1) \cdot p(celebrations|c1)\right]$

$\therefore p(c1|doc) \propto \frac{3}{5} \cdot \frac{(2+1)}{(7+4)} \cdot \frac{(2+1)}{(7+4)} \cdot \frac{(2+1)}{(7+4)} \cdot \frac{(0+1)}{(7+4)}$

$\therefore p(c1|doc) \propto \frac{3}{5} \cdot \frac{3}{11} \cdot \frac{3}{11} \cdot \frac{3}{11} \cdot \frac{1}{11} = 0.0011$

Now, similarly,

$$P(c_2|doc) \propto P(c_2) \cdot P(happy|c_2) \cdot p(new|c_2) \cdot p(year|c_2) \cdot p(celebrations|c_2)$$

$$\Rightarrow P(c_2|doc) \propto \frac{2}{5} \cdot \frac{(0+1)}{(4+3)} \cdot \frac{(0+1)}{(4+3)} \cdot \frac{(0+1)}{(4+3)} \cdot \frac{(0+1)}{(4+3)}$$

$$= 0.00016$$

since $P(c_1|doc) > P(c_2|doc)$, the given document is assigned to "c1".

Task-2] (2.1) $p(SPAM|x_1) = \dfrac{\exp \sum \lambda_i f_i(SPAM, x)}{\sum_{y'} \exp \sum \lambda_i f_i(y, x)}$

Now, $\sum_i \lambda_i f_i(SPAM, x) = \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3 + \lambda_4 f_4 + \lambda_5 f_5 + \lambda_6 f_6 + \lambda_7 f_7 + \lambda_8 f_8$

$= \lambda_1 f_1 + \lambda_3 f_3 + \lambda_5 f_5 + \lambda_7 f_7$

$= (0.2)(1) + (0.5)(1) + (0) + (0.1)(1)$

$= 0.8$

similarly,

$$\sum_i \lambda_i f_i(\text{HAM}, x_1) = \lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3 + \lambda_4 f_4 + \lambda_5 f_5 + \lambda_6 f_6 + \lambda_7 f_7 + \lambda_8 f_8.$$

$$= (-0.1)(1) + (-0.2)(1) + 0 + 0.$$

$$= -0.3.$$

$$\therefore p(\text{SPAM}|x_1) = \frac{\exp(0.8)}{\exp(0.8) + \exp(-0.3)} = \frac{2.2255}{2.9663}$$

$$\therefore P(\text{SPAM}|x_1) = 0.7502.$$

subtask: 2.2 Log-likelihood

$$\log P_\lambda(Y|X) = \sum_{(y,x) \in (Y,X)} \log P_\lambda(Y|X).$$

$$= \sum_{(y,x) \in (Y,X)} \log \frac{\exp \sum_i \lambda_i f_i(y,x)}{\sum_{y'} \exp \sum_i \lambda_i f_i(y',x)}.$$

$$= \sum_{(y,x)} \log \exp \sum_i \lambda_i f_i(y,x) - \sum_{(y,x)} \log \sum_{y'} \exp \sum_i \lambda_i f_i(y',x)$$

$$\therefore \log P_\lambda(Y|X) = A_\lambda - B_\lambda.$$

Now, partial derivatives :-

$$\frac{\partial A_\lambda}{\partial \lambda_6} = \sum_{(y,x)\in(Y,X)} f_6(y,x) = 1+1 = 2.$$

also, $\frac{\partial B_\lambda}{\partial \lambda_6} = P_\lambda(HAM \mid x_3) \cdot f_6(HAM, x_3) + P_\lambda(HAM \mid x_5) \cdot f_6(HAM, x_5)$

$$+ P_\lambda(HAM \mid x_6) \cdot f_6(HAM, x_6).$$

$$\therefore \frac{\partial B_\lambda}{\partial \lambda_6} = P_\lambda(HAM \mid x_5) + P_\lambda(HAM \mid x_6).$$

Now, $P_\lambda(HAM \mid x_5) = \frac{\exp(-0.2+0.4)}{\exp(0.2)+\exp(0.5)} = \frac{1.2214}{2.87} = 0.425$

and $P_\lambda(HAM \mid x_6) = \frac{\exp(0.3)}{\exp(0.3)+\exp(0.2)} = \frac{1.3498}{2.5712} = 0.525$

$$\therefore \frac{\partial B_\lambda}{\partial \lambda_6} = 0.425 + 0.525 = 0.95.$$

$$\therefore \frac{\partial A_\lambda}{\partial \lambda_6} - \frac{\partial B_\lambda}{\partial \lambda_6} = 2 - 0.95 = 1.05$$

Task: 3] kNN - classification

For $k = 1$; cross will be assigned to "empty circle" with prob = 1.

For $k = 3$; cross will be assigned to "empty circle" with prob = $2/3$.

For $k = 5$; cross will be assigned to "empty circle" with prob = $3/5$.

[Please Note: o → empty circle
● → filled circle]

⇒ Hence, For $k = 5$, the kNN - classifier obtains the lowest classification confidence for the cross. //