Cheng Chi Fung 12219691

Q1. Probabilities Computation

(i) $P(B_1 = 1) = (1/3)$

(ii) $Likelihood(B_2 = 0) = P(B_2 = 0 \,|B_1 = 1) = 1$

(iii) $P(B_1 = 1 |\, B_2 = 0) = \, 1/3$

(iv) Yes, I will change my choice.

$$= \, P(B_1 = 1 |\, B_2) = \frac{P(B_1 = 1 \mid B_1) P(B_1)}{P(B_2 |\, B_1) P(B_1)} = \frac{(1/2)(1/3)}{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)(1)} = (1/3)$$

$$= P(B_3 = 1 |\, B_2) = \frac{P(B_3 = 1 |\, B_1) P(B_1)}{P(B_2 |\, B_1) P(B_1)} = \frac{(1)(1/3)}{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + (0)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)(1)} = (2/3)$$

Q2. Naïve Bayes Classifier

(i) Let $P(y = i) = \theta_k$

likelihood = $P(N_{doc} | \theta_k) = {\theta_1}^{N_1} {\theta_2}^{N_2} \ldots {\theta_K}^{N_K} = \prod_{i=1}^{K} {\theta_i}^{N_i}$

log-likelihood = $\log\left(\prod_{i=1}^{K} {\theta_k}^{N_i}\right) = \sum_{i=1}^{K} N_i \log(\theta_k)$

max(log-likelihood) = $f(\theta) = \arg\max_{\theta_k} \sum_{i=1}^{K} N_i \log(\theta_k)$

$$\text{s.t. } g(\theta) = \sum_{i=1}^{K} \theta_k = 1$$

Solve it by langrage multiplier,

$$\frac{\partial L}{\partial \theta_k} = \frac{\partial f}{\partial \theta_k} - \frac{\partial g}{\partial \theta_k} = N_i \log(\theta_k) \frac{1}{\ln(\theta_k)} - \lambda \theta_k = 0$$

$$\lambda \theta_k = \frac{N_i \log(\theta_k)}{\ln(\theta_k)}$$

$$\lambda \theta_k = N_i$$

Since, $\sum_{i=1}^{K} \theta_k = 1$, so $\lambda = N_{doc}$, so $N_{doc}\theta_k = N_i$, $\theta_k = \frac{N_i}{N_{doc}}$

Max likelihood estimate probability of $\theta_k = \frac{N_k}{N_{doc}}$

(ii) Let $P(w_i \mid y = k) = \theta_{k,i}$

Max likelihood estimate probability of $\theta_{k,i} = \frac{count(w_i,k)}{\sum_j^V count(w_j,k)}$

(iii) We can use Laplace Smoothing in $P(w_i \mid y = k)$ to avoid zero probability.

This can be done by changing the calculation of $P(w_i \mid y = k) = \frac{count(w_i,k)}{\sum_j^V count(w_j,k)}$

to $\frac{count(w_i,k)+\lambda}{\sum_j^V count(w_j,k)+\lambda\times|V|}$, which $1 \geq \lambda > 0$

(iv) We can simply ignore the words only occur in test set.

We can utilize stop words corpus or some corpus containing very frequent words to filter out those words before the construction of the bags of words.

Q3. Running the Classifier on an Example

(i) $P(y=+) = 0.625$, $P(y=-) = 0.375$

(ii) $P(w=easy|y=+) = 0.08333333333333333$

$P(w=easy|y=-) = 0$

$P(w=very|y=+) = 0$

$P(w=very|y=-) = 0.125$

$P(w=great|y=+) = 0.08333333333333333$

$P(w=great|y=-) = 0$

(iii) $P(y=+) = 0.625$, $P(y=-) = 0.375$

$P(w=easy|y=+) = 0.06896551724137931$

$P(w=easy|y=-) = 0.04$

$P(w=very|y=+) =0.034482758620689655$

$P(w=very|y=-) =0.08$

$P(w=great|y=+) =0.06896551724137931$

$P(w=great|y=-) = 0.04$

(iv) $P\ (y = +|d) = 1.02505228e\text{-}04$

$P\ (y = -|d) = 4.80000000e\text{-}05$

The final result will be positive