Text Categorization with Naïve Bayes Classifier

$d = x_1 x_2 \ldots x_L$ where $x_i \in V$

IF θ_i represents category i accurately, then…

How can we make this happen?

$$\text{category}(d) = \arg\max_i p(\theta_i \mid d)$$

$$= \arg\max_i p(d \mid \theta_i) p(\theta_i)$$

$$= \arg\max_i \prod_{w \in V} p(w \mid \theta_i)^{c(w,d)} p(\theta_i)$$

$p(w \mid \theta_1)$

d ←

sports	0.02
game	0.01
basketball	0.005
football	0.004
…	

$p(\theta_1)$

$p(w \mid \theta_2)$

d ←

travel	0.05
attraction	0.03
trip	0.01
…	

$p(\theta_2)$

$p(w \mid \theta_k)$

d ←

science	0.04
scientist	0.03
spaceship	0.006
…	

$p(\theta_k)$

$$\text{category}(d) = \arg\max_i \log p(\theta_i) + \sum_{w \in V} c(w,d) \log p(w \mid \theta_i)$$