1 Arun Moosame

$t \sim Categorical(\pi_1 \dots \pi_k)$

$x_i \mid t = k \sim Bernoulli(\theta_{k,i})$

$$p(t \mid x) = \frac{p(x \mid t)\, p(t)}{p(x)}$$

$t=1$ Spam $t=0$ Not spam

$x_i \in \{0,1\}$ means ith word in vocabulary occurs in document

$$\propto p(t)\, p(x \mid t)$$

$p(\text{spam} \mid \text{words})$

$$= p(t=1 \mid x_1 \dots\dots x_m) = \frac{p(x^{(1)} \dots\dots x^{(m)} \mid t=1)\, p(t=1)}{p(x^{(1)} \dots x^{(m)})}$$

$$\propto p(t=1)\, p(x^{(1)} \dots x^{(m)} \mid t=1)$$

Naive assumption: each word independent of other words

$$= p(t=1)\, p(x^{(1)} \mid t=1)\, p(x^{(2)} \mid t=1) \dots\dots p(x^{(m)} \mid t=1)$$

$$= p(t=1) \prod_{m=1}^{M} p(x^{(m)} \mid t=1)$$

2

$t \sim \text{Categorical}(\pi_1, \dots, \pi_K)$

$x_i \mid t = k \sim \text{Bernoulli}(p_{k,i})$

$\hat{\pi}, \hat{p}_{k,i}$

$$\arg\max \log p(X, t)$$

$$\arg\max \sum_{m=1}^{M} \log p(x^{(m)}, t^{(m)})$$

$$\arg\max \sum_{m=1}^{M} \log p(t^{(m)}) p(x^{(m)} \mid t^{(m)})$$

$$\arg\max \sum_{m=1}^{M} \log p(t^{(m)}) + \log p(x^{(m)} \mid t^{(m)})$$

$$\arg\max \sum_{m=1}^{M} \log p(t^{(m)}; \pi) + \sum_{m=1}^{M} \log p(x^{(m)} \mid t^{(m)}; p_{k,i})$$

$$\arg\max \sum_{m=1}^{M} \log p(t^{(m)}; \pi) + \sum_{k=1}^{K} \sum_{\substack{t^{(m)}=k \\ m=1 \dots M}} p(x^{(m)} \mid t^{(m)} = k; p_{k,i})$$

just counting

$$\hat{\pi}_k = \frac{\sum_{m=1}^{M} \mathbb{1}\{t^{(m)} = k\}}{M} \quad \text{for } k = 1 \dots K$$

MLE for Multinomial distribution (covered in class notes)