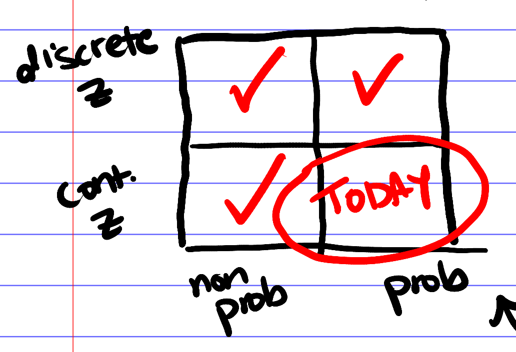
CS181: Topic Models



Also:
- HW5 is out
- practical out on Friday

↖ unsupervised learning

Book Ch. 9.6

(z) hidden (continuous) → (x) observed

Goal: Learn a model $z \to x$

Many versions/examples:

① Factor Analysis

$A \sim N(0, I)$ size $D \times K$

for every n: $z_n \sim N(0, I)$ size K

$x_n = Az_n + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$

Note: • if we knew z, then finding A is relatively easy → regression!
• if we knew A, then solving for z is relatively easy

$p(z|x,A) \propto p(x|z,A) \cdot p(z)$

↳ algorithm that alternates btw opt z|A, opt A|z

② Variational Autoencoder

$z \sim N(0, I)$ (size K)

$x = f_\theta(z) + \varepsilon$, $\varepsilon \sim N(0, I)$

f_θ: neural net char. by θ; θ are the global params

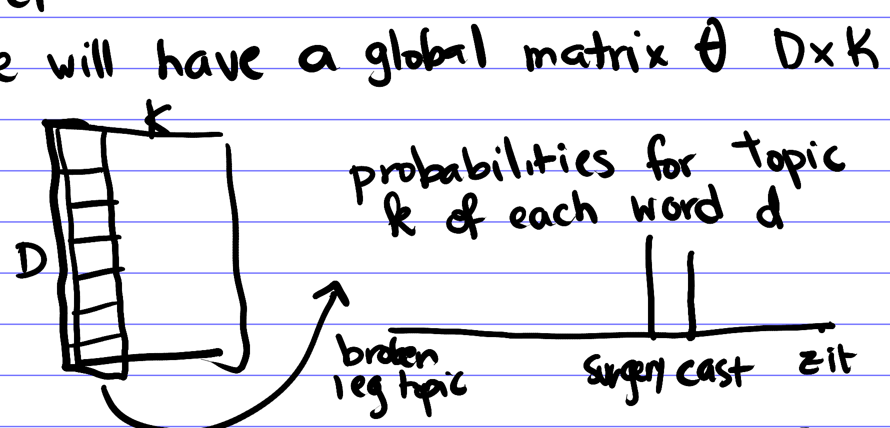
learn an inference network:

$p(z|x,\theta) \approx q_\phi(z)$

$p(z|x,\theta)$: now hard to compute; q_ϕ: new neural net with params ϕ

③ Deep Dive: Topic Models

Model:
- We will have a global matrix θ $D \times K$

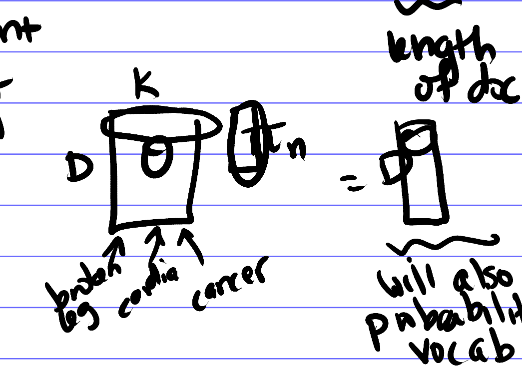


- For each document $\pi_n \sim Dir(\alpha)$

LOCAL HIDDEN: this is a process to generate a vector of probabilities size K (1 ... K)

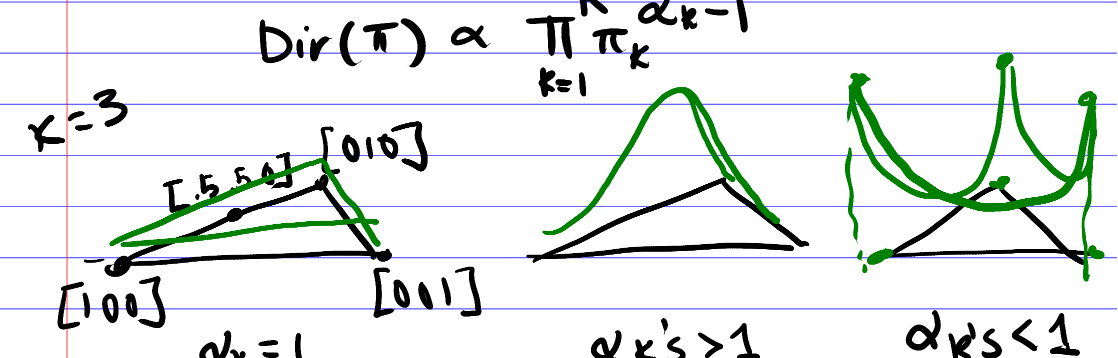
$x_n \sim Mult(\theta \pi_n, L)$

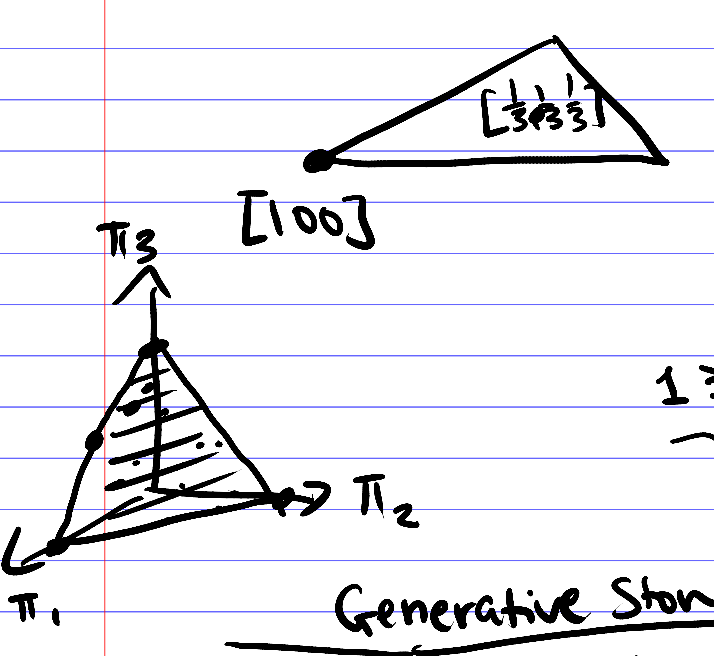
x_n: document counts of each word in the doc (size D); L: length of doc



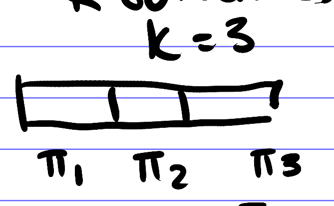
What is a Dirichlet?

$Dir(\pi) \propto \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$





all valid prob dist on K outcomes, K=3



$1 \geq \pi_k \geq 0$, $\sum \pi_k = 1$

Generative Story ↑
Inference ↓

Objective: Log-Lik of the model:

$$p(x_n|\pi_n,\theta) = \prod_d \left(\sum_k \pi_{nk}\theta_{kd}\right)^{x_{nd}}$$

$\prod_d$: prod over words

$$p(\pi_n) \propto \prod_k \pi_{nk}^{\alpha_k - 1} \quad \text{(dims of x)}$$

$$p(x,\pi|\theta,\alpha) \propto \left(\prod_k \pi_{nk}^{\alpha_k-1}\right)\left(\prod_d \left(\sum_k \pi_{nk}\theta_{kd}\right)^{x_{nd}}\right)$$

note: inference is over just θ and $\{\pi_n\text{'s}\}$

this is ugly for inference!

Approach: Introduce an aux. variable:

Equiv generative model:

Given θ (parameter)

For every doc n:

$\pi_n \sim Dir(\alpha)$

For each of the L words:

$z_{nl} \sim$ ~~Mult~~ Cat(π_n) } what topic does word l come from

$w_{nl} \sim$ ~~Mult~~ Cat$(\theta_{z_{nl}})$

and x_n is the set of counts of each word.

$$p(x_n|\{z\}_n,\pi_n,\theta) = \prod_l \prod_d \prod_k \theta_{dk}^{\mathbb{I}(w_{nl}=d)\cdot\mathbb{I}(z_{nl}=k)}$$

$$p(\{z\}_n|\pi_n) = \prod_l \prod_k \pi_{nk}^{\mathbb{I}(z_{nl}=k)}$$

$$p(\pi_n) \propto \prod_k \pi_{nk}^{\alpha_k - 1}$$

Joint Prob is the product of these terms

$$\log p(\{x,\pi,z\}|\theta,\alpha) \propto$$

$$\sum_n \left[\sum_k \left(\alpha_k - 1 + \sum_l \mathbb{I}(z_{nl}=k)\right)\log \pi_{nk} + \sum_d \sum_k \left(\sum_l \mathbb{I}(w_{nl}=d)\,\mathbb{I}(z_{nl}=k)\right)\log \theta_{dk}\right]$$

$\mathbb{I}(w_{nl}=d)$: known

Can now apply E.M.!

$$\mathbb{E}_z\left[\log p(\{x,z,\pi\}|\theta)\right]$$

$$= \sum_n \left[\sum_k \left(\alpha_k - 1 + \sum_l q_{nlk}\right)\log \pi_{nk} + \sum_d \sum_k \left(\sum_l q_{nlk}\,\mathbb{I}(w_{nl}=d)\right)\log \theta_{dk}\right]$$

q_{nlk}: partial/soft assignments of topic to doc.

q_{nlk} is prob $z_{nl} = k$

$$p(z_{nl}|\pi_n, w_{nl}, \theta) \propto \pi_{nk}\,\theta_{dk} = q_{nlk}$$

π_{nk}: prior on $z_{nl}|\pi_n$; θ_{dk}: likelihood term