Lecture 11 Manipulating Moments

Wednesday, September 28, 2016 10:32 PM

- Recap: Tensor Decomposition can find $\{\lambda_i, a_i\}$

given $$T = \sum_{i=1}^{r} \lambda_i \, a_i \otimes a_i \otimes a_i$$

$$M = \sum_{i=1}^{r} \lambda_i a_i a_i^T$$

- Note: a_i's may not have norm 1, but we can still compute both λ_i, a_i (no scale symmetry)

This is because after whitening, $T(w,W,w) = \sum_{i=1}^{r} \underbrace{\lambda_i^{-\frac{1}{2}}}_{\text{can get } \lambda_i} (\underbrace{\sqrt{\lambda_i}\, W^T a_i}_{\text{unit vector}})^{\otimes 3}$

- cannot get $\lambda_i / \|a_i\|$ if we don't have M, and use Jenrich.

- Problem: how to get M and T.

- Example 1: mixture of spherical Gaussians

Model: r Gaussians, each with mean $\mu_i \in \mathbb{R}^d$, variance $\sigma^2 I$.

with probability p_i $x \sim N(\mu_i, \sigma^2 I)$

- Moments: $$\mathbb{E}[x] = \sum_{i=1}^{r} p_i \mu_i$$

$$\mathbb{E}[xx^T] = \sum_{i=1}^{r} p_i \mu_i \mu_i^T + \sigma^2 I$$

$$\mathbb{E}[x \otimes x \otimes x]$$

$$= \mathbb{E}[(\underbrace{\mu}_{\substack{=\mu_i \\ \text{w.p. } p_i}} + \underbrace{\eta}_{\text{Gaussian noise } \sigma^2 I})^{\otimes 3}]$$

$$= \mathbb{E}[\mu^{\otimes 3}] + \mathbb{E}[\mu \otimes \eta \otimes \eta] + \cdots \text{(symmetric)}$$

$$\boxed{+ \mathbb{E}[\mu \otimes \mu \otimes \eta] + \cdots \text{(symmetric)}}$$

$$+\mathbb{E}[\mu\otimes\mu\otimes\eta]+\cdots$$
$$\underbrace{+\mathbb{E}[\eta\otimes\eta\otimes\eta]}_{=0 \text{ because } \eta\sim N(0,\sigma^2 I)}$$

$$=\sum_{i=1}^{r} p_i \mu_i \mu_i^T + T_{extra}$$

$$T_{extra} = \mu\otimes\sigma^2 I + \text{symmetric forms}$$

$$T_{extra}[i,j,k] = \sigma^2\left(\mathbb{E}[\mu]_i \delta_{jk} + \mathbb{E}[\mu]_j \delta_{ik} + \mathbb{E}[\mu]_k \delta_{ij}\right)$$
$$(\delta_{ij}=1 \text{ iff } i=j)$$

- Problem: extra $\sigma^2 I$ in $\mathbb{E}[xx^T]$
  extra T_{extra} in $\mathbb{E}[x\otimes x\otimes x]$

Finding σ^2: $\sum_{i=1}^{r} p_i \mu_i \mu_i^T$ rank r

so $\sigma^2 = \sigma_{r+1}(\mathbb{E}[xx^T])$

Handling T_{extra}: T_{extra} can be computed given $\mathbb{E}[x]=\mathbb{E}[\mu]$ and σ^2 !

- Example 2: Latent Dirichlet Allocation

recall: model: each document talks about a mixture of topics, mixing weight

$$w\sim \mathrm{Dir}(\alpha)$$
$$\Pr[w] \propto \prod_{i=1}^{r} w_i^{\alpha_i - 1}, \quad \alpha_0 = \sum_{i=1}^{r}\alpha_i$$

Fact: $\mathbb{E}[w_i] = \dfrac{\alpha_i}{\alpha_0}$

$$\mathbb{E}[w_i w_j] = \frac{\alpha_i\alpha_j}{\alpha_0(\alpha_0+1)} \qquad \mathbb{E}[w_i^2] = \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)}$$

$$\mathbb{E}[w_i w_j] = \frac{\alpha_i \alpha_j}{\alpha_0(\alpha_0+1)} \quad \mathbb{E}[w_i^2] = \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)}$$

$$\mathbb{E}[w_i w_j w_k] = \frac{\alpha_i \alpha_j \alpha_k}{\alpha_0(\alpha_0+1)(\alpha_0+2)} \quad \mathbb{E}[w_i^3] = \frac{\alpha_i(\alpha_i+1)(\alpha_i+2)}{\alpha_0(\alpha_0+1)(\alpha_0+2)}$$

Moments:

first moment : $\Pr[\text{1st word} = i] = \sum_{l=1}^{r} A_{i,l} \cdot \mathbb{E}[w_l]$

$$u = \sum_{l=1}^{r} \frac{\alpha_l}{\alpha_0} a_l$$

second moment: $\Pr[\text{1st word}=i, \text{2nd word}=j]$

recall $\quad Q = ARA^T \quad (R = \mathbb{E}[ww^T])$

$$\mathbb{E}[ww^T] = \frac{\mathrm{Diag}(\alpha_i)}{\alpha_0(\alpha_0+1)} + \frac{\alpha\alpha^T}{\alpha_0(\alpha_0+1)}$$

So $\quad Q = \sum_{l=1}^{r} \frac{\alpha_l}{\alpha_0(\alpha_0+1)} a_l a_l^T + \frac{1}{\alpha_0(\alpha_0+1)} (\alpha_0 u)(\alpha_0 u)^T$

if we know α_0, can get rid of second term!

works similarly for third order tensor.