Lecture 6 Spectral Methods: Tensor Methods

Recap on Gaussian Mixtures.

$$A = [\vec{u}_1 \mid \vec{u}_2 \mid \dots \mid \vec{u}_k]_{d\times k}$$

$$x = A\vec{h} + z$$

h: latent (hidden) variable, it is not observed.

$$h \in \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_k\} \qquad \mathbb{E}[h] = w.$$

↳ selects a column from A

$$z \sim \mathcal{N}(0, \sigma^2 I)$$

Method of Moments: Estimate emprical moments from the data.

Gaussian Approx → first and second order moments are sufficient

first ↓ mean; second order ↓ covariance

MLE: $\max_{\theta} \sum_{i=1}^{n} \log p_\theta(x_i)$ What is the challenge for Gaussian Mixtures?

⇒ Non-convex

Expectation Maximization → Approximation to MLE

local optimum

w & A & σ^2 are the unknown parameters.

MoM: $P(\vec{h} = \vec{e}_i) = (\mathbb{E}[\vec{h}])_i = (\vec{w})_i$ where $\mathbb{E}[\vec{h}] = \vec{w}$

Moments: $\mathbb{E}[\vec{x}] = \mu = \sum_i w_i \mu_i = Aw$

Correlation ⇒ $\mathbb{E}[xx^T] = \mathbb{E}[(Ah+z)(Ah+z)^T] = A\,\mathbb{E}[hh^T]A^T + A\,\mathbb{E}[hz^T]$

$(h \perp z \rightarrow$ independent $\qquad + \mathbb{E}[hz]A^T + \mathbb{E}[zz^T]$

$$\mathbb{E}[h]\underset{\substack{\downarrow \\ 0}}{\mathbb{E}[z]}A^T \quad \sigma^2 I$$

$$= A\,\mathbb{E}[hh^T]\,A^T + \sigma^2 I$$

$$\mathbb{E}[hh^T] = \sum_{i=1}^{n} w_i e_i e_i^T$$

$$= \mathrm{Diag}(\vec{w})$$

$$\mathbb{E}[xx^T] = A\,\mathrm{Diag}(\vec{w})A^T + \underbrace{\sigma^2 I}_{\text{if we know it, omit it}}$$

Can we recover A from $A\,\mathrm{Diag}(\vec{w})\,A^T = \mathbb{E}[xx^T]$

If A is orthogonal, $w \neq 0$, $\begin{pmatrix} w_i \neq w_j \\ i \neq j \end{pmatrix}$, A is unique.

$\mathrm{Span}(A) \rightarrow$ eigenspace

Assume $k \ll d \rightarrow$ low-rank approximation is useful.

Learning Gaussian mixtures through clustering

Learning A through clustering.

- $\rightarrow$ Project x to $\mathrm{span}(A)$
- $\rightarrow$ Distance-based clustering (k-means)

Failure to cluster under high variance.

Tensor Notation for higher order moments

$$\mathbb{E}[x \otimes x] = \mathbb{E}[xx^T] \in \mathbb{R}^{d \times d} \rightarrow \text{second order tensor}$$

$$\mathbb{E}[x \otimes x \otimes x] \in \mathbb{R}^{d \times d \times d} \rightarrow \text{third order tensor}$$

$$\mathbb{E}[x \otimes x \otimes x]_{i_1 i_2 i_3} = \mathbb{E}[x_{i_1} x_{i_2} x_{i_3}]$$

For Gaussian mixture,

$$\mathbb{E}[(Ah) \otimes Ah \otimes Ah]$$

$$\mathbb{E}[h \otimes h \otimes h] = \sum_{i=1}^{k} w_i \ \vec{e_i} \otimes \vec{e_i} \otimes \vec{e_i}$$

$$\mathbb{E}[Ah \otimes Ah] = A\,\mathrm{diag}(\vec{w})\,A^T = \sum_{i=1}^{k} w_i \mu_i \mu_i^T$$

$$\mathbb{E}[Ah \otimes Ah \otimes Ah] = \sum_{i=1}^{k} w_i \ \underbrace{\vec{\mu_i} \otimes \vec{\mu_i} \otimes \vec{\mu_i}}_{\text{rank } 1}$$

(C.P.) Tensor Decomposition: Identifiability conditions

Canonicaldecomposition

Parafac

(symmetric)

Rank-1 matrix : $(uu^T)_{ij} = u_i u_j$ → $u \otimes u$

Rank-1 Tensor: $(u \otimes u \otimes u)_{ijk} = u_i u_j u_k$

CP-Rank can be larger than d.

NP-hard to compute

$T \in \mathbb{R}^{d \times d \times d}$ suppose symmetric.

$$\left(\begin{array}{l} Mv = \sum m_i v_i \\ M = [\vec{m_1} | \dots | \vec{m_d}] \end{array} \right)$$

$T_{i,j,k}$ — mode-1, mode2, mode-2

Contraction operation: $T(\bar{u}, \cdot, \cdot)$

$$= \sum_i T_{i,\cdot,\cdot} \, u_i$$

u_i: can be a matrix as well

$T_{i,\cdot,\cdot}$: matrix in i^{th} slice

$$\{T(A,B,C)\}_{i',j',k'} = \sum_{i,j,k} T_{i,j,k} A_{ii'} B_{jj'} C_{kk'}$$

We want to $\max_{\|u\|=1} T(u,u,u)$ $\left(\max_{\|u\|=1} u^T M u \text{ matrix case} \right)$

$$\sum T_{i,j,k} u_i u_j u_k$$

Then, writing the Lagrangian, $\longrightarrow \sum_{i,j,k} T_{i,j,k}\, u_i u_j u_k$

$$\mathcal{L}(\lambda) = \max_{u}\; T(u,u,u) - \frac{3}{2}\lambda(u^T u - 1)$$

First Order Condition: $3\,T(u,u,.) = 3\lambda u$

$$\underbrace{T(u,u,.) = \lambda u}_{\text{Quadratic in many direction}} \longrightarrow \text{eigenvector}$$

$Mu = \lambda u$ is linear

For linear system, you can have d eigenvector.

For tensors, you have exponentially many eigenvectors.

Second Order Conditions: $v^T(M - \lambda I)v < 0 \quad v \perp u$

Hard for tensors

Special Case: $T = \sum \lambda_i\, \vec{u_i} \otimes \vec{u_i} \otimes \vec{u_i} \quad (u_i \perp u_j \quad i \neq j)$

$\hookrightarrow$ orthogonal tensor

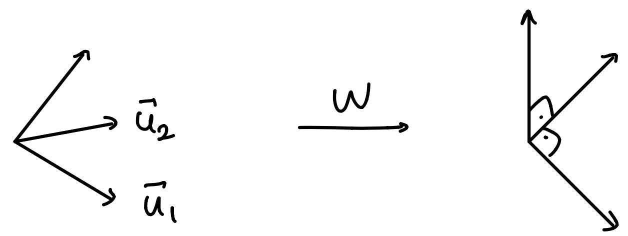
* For orthogonal tensors, local optima $\{u_i\}$.
(stable point)

$\longrightarrow$ How can we solve it for Gaussian Mixtures ?

Power Method: (orthogonal Tensors) $u \leftarrow \dfrac{T(u,u,.)}{\| T(u,u,.)\|}$ Stability result

$\longrightarrow u_i$ gives u_i

Converging to local optima is guaranteed.



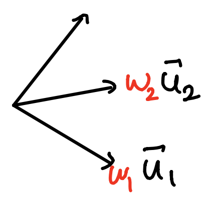
orthogonal

$$W^T A = R$$

Whitening Transformation

$$(W^T A)(A^T W) = I$$

$A \, \mathrm{Diag}(\vec{w}) A^T$ (scaled u_i's)



How to use

↳ Whitening Transform?

$$W^T (A \, \mathrm{Diag}(\vec{w}) A^T) W = I$$

* Can we go back from whitened vectors?

If vectors were linearly independent, yes!