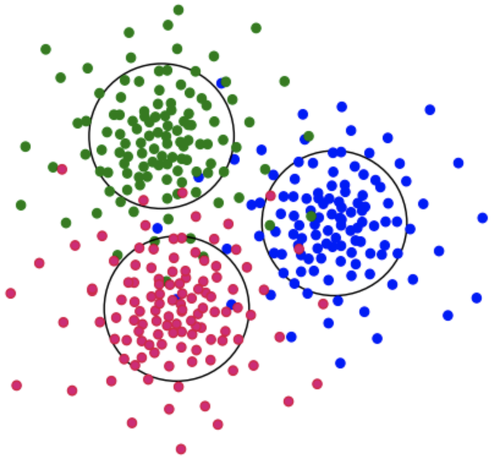
Lecture 11 CMS 165

Spectral Methods: Tensor methods

Recap: PCA on Gaussian Mixtures

- k Gaussians: each sample is $x = Ah + z$.
- $h \in [e_1, \ldots, e_k]$, the basis vectors. $\mathbb{E}[h] = w$.
- $A \in \mathbb{R}^{d \times k}$: columns are component means.
- Let $\mu := Aw$ be the mean.
- $z \sim \mathcal{N}(0, \sigma^2 I)$ is white Gaussian noise.



$$\boxed{\mathbb{E}[(x-\mu)(x-\mu)^\top] = \sum_{i \in [k]} w_i (a_i - \mu)(a_i - \mu)^\top + \sigma^2 I.}$$

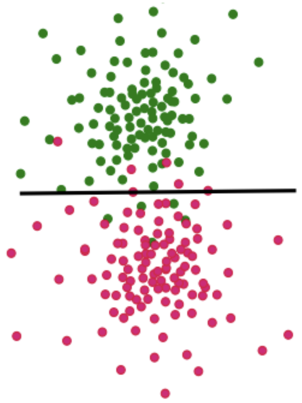
Can obtain span(A).

But what about columns of A?

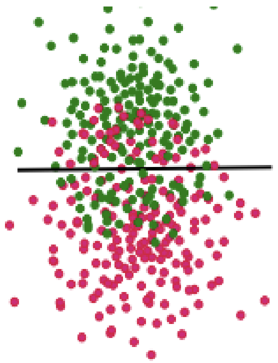
Learning Gaussian mixtures through clustering

Learning A through Spectral Clustering

- Project samples x on to span(A).
- Distance-based clustering (e.g. k-means).
- A series of works, e.g. Vempala & Wang.



Failure to cluster under large variance.



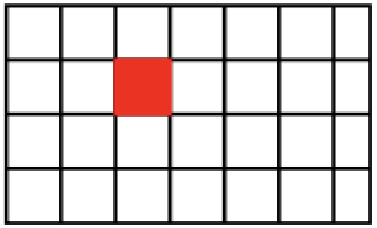
Learning Gaussian Mixtures Without Separation Constraints?

Tensor Notation for Higher Order Moments

- Multi-variate higher order moments form tensors.
- Are there spectral operations on tensors akin to PCA?

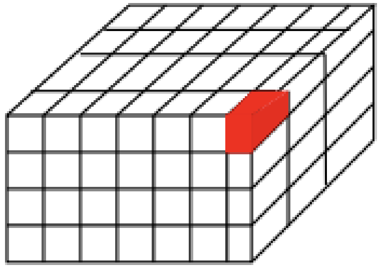
Matrix

- $\mathbb{E}[x \otimes x] \in \mathbb{R}^{d \times d}$ is a second order tensor.
- $\mathbb{E}[x \otimes x]_{i_1,i_2} = \mathbb{E}[x_{i_1} x_{i_2}]$.
- For matrices: $\mathbb{E}[x \otimes x] = \mathbb{E}[x x^\top]$.



Tensor

- $\mathbb{E}[x \otimes x \otimes x] \in \mathbb{R}^{d \times d \times d}$ is a third order tensor.
- $\mathbb{E}[x \otimes x \otimes x]_{i_1,i_2,i_3} = \mathbb{E}[x_{i_1} x_{i_2} x_{i_3}]$.



Third order moment for Gaussian mixtures

- Consider mixture of k Gaussians: each sample is $x = Ah + z$.
- $h \in [e_1, \ldots, e_k]$, the basis vectors. $\mathbb{E}[h] = w$.
- $A \in \mathbb{R}^{d \times k}$: columns are component means. $\mu := Aw$ be the mean.
- $z \sim \mathcal{N}(0, \sigma^2 I)$ is white Gaussian noise.

$$\mathbb{E}[x \otimes x \otimes x] = \sum_i w_i a_i \otimes a_i \otimes a_i + \sigma^2 \sum_i (\mu \otimes e_i \otimes e_i + \ldots)$$

Intuition behind equation

$$\begin{aligned}\mathbb{E}[x \otimes x \otimes x] &= \mathbb{E}[(Ah) \otimes (Ah) \otimes (Ah)] + \mathbb{E}[(Ah) \otimes z \otimes z] + \ldots \\ &= \sum_i w_i \cdot a_i \otimes a_i \otimes a_i + \sigma^2 \sum_i \mu \otimes e_i \otimes e_i + \ldots\end{aligned}$$

How to recover parameters A and w from third order moment?

Tensor Slices

$$M_3 = \sum_i w_i a_i \otimes a_i \otimes a_i, \quad M_2 = \sum_i w_i a_i \otimes a_i.$$

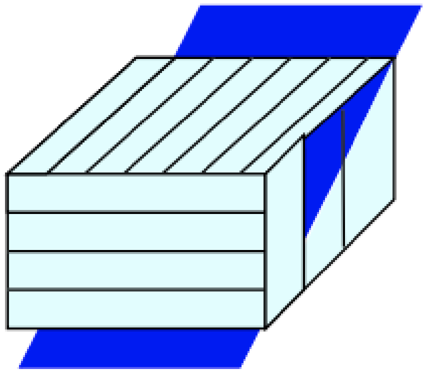
Multilinear transformation of tensor

$$\boxed{M_3(B, C, D) := \sum_i w_i (B^\top a_i) \cdot (C^\top a_i) \cdot (D^\top a_i)}$$

Slice of a tensor

$M_3(I, I, r) = \sum_i w_i a_i \otimes a_i \langle a_i, r \rangle = A\mathrm{Diag}(w)\mathrm{Diag}(A^\top r)A^\top$

$$\begin{aligned} M_3(I, I, r) &= A \cdot \mathrm{Diag}(w)\mathrm{Diag}(A^\top r) \cdot A^\top \\ M_2 &= A \cdot \mathrm{Diag}(w) \cdot A^\top \end{aligned}$$



Eigen-decomposition

$$\boxed{M_3(I,I,r) = A \cdot \mathrm{Diag}(w)\mathrm{Diag}(A^\top r) \cdot A^\top, \quad M_2 = A \cdot \mathrm{Diag}(w) \cdot A^\top.}$$

Assumption: $A \in \mathbb{R}^{d\times k}$ has full column rank.

- $M_2 = U\Lambda U^\top$ be eigen-decomposition. $U \in \mathbb{R}^{d\times k}$.
- $U^\top M_2 U = \Lambda \in \mathbb{R}^{k\times k}$ is invertible.

$$\boxed{X = \left(U^\top M_3(I,I,r)U\right)\left(U^\top M_2 U\right)^{-1} = V \cdot \mathrm{Diag}(\tilde{\lambda}) \cdot V^{-1}.}$$

- Substitution: $X = (U^\top A)\mathrm{Diag}(A^\top r)(U^\top A)^{-1}$.
- We have $\boxed{v_i \propto U^\top a_i}$.

Technical Detail

$r = U\theta$ and θ drawn uniformly from sphere to ensure eigen gap.

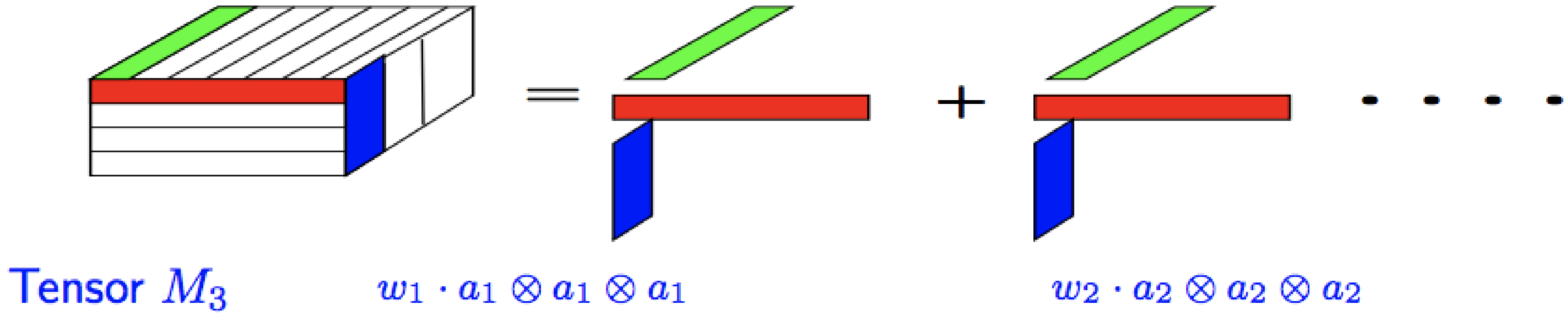
Shortcomings

- The resulting product is not symmetric. Eigen-decomposition $V\text{Diag}(\tilde{\lambda})V^{-1}$ does not result in orthonormal V. More involved in practice.
- Require good eigen-gap in $\text{Diag}(\tilde{\lambda})$ for recovery. For $r = U\theta$, where θ is drawn uniformly from unit sphere, gap is $1/k^{2.5}$. Numerical instability in practice.
- $M_3(I, I, r)$ is only a (random) slice of the tensor. Full information is not utilized.

Tensor Factorization

- Recover A and w from M_2 and M_3.

$$M_2 = \sum_i w_i a_i \otimes a_i, \quad M_3 = \sum_i w_i a_i \otimes a_i \otimes a_i.$$



- $a \otimes a \otimes a$ is a rank-1 tensor since, its $(i_1, i_2, i_3)^{\text{th}}$ entry is $a_{i_1} a_{i_2} a_{i_3}$.
- M_3 is a sum of rank-1 terms.
- When is it the most compact representation? (Identifiability).
- Can we recover the decomposition? (Algorithm?)

References

- Monograph on spectral learning on matrices and tensors (preprint available on Piazza)
- Slides from MLSS: available on http://tensorlab.cms.caltech.edu/users/anima/