Remarks on Tensor power Method.

1) about the iterations in practice you are not give

$$T = \sum_{i=1}^{k} \lambda_i \, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$$

but you are given $T^{\alpha\beta\gamma}$ set of numbers cube of size $D \times D \times D$.

$$x^t = \frac{T(I, x^{t-1}, x^{t-1})}{\| T(I, x^{t-1}, x^{t-1}) \|_2}$$

$$\boxed{\left[T(I, x^{t-1}, x^{t-1}) \right]^\alpha} = \sum_{\beta\gamma} T^{\alpha\beta\gamma} (x^{t-1})^\beta (x^{t-1})^\gamma$$

$$\Updownarrow$$

exercise check $T(I, x^{t-1}, x^{t-1}) = \sum_i \lambda_i \vec{v}_i (\vec{v}_i^T \cdot x^{t-1})^2$

2) you can't know of the condition

$$|\lambda_1 \vec{v}_1^T \cdot \vec{x}^0| > |\lambda_2 \vec{v}_2^T \cdot \vec{x}^0|$$

but you just iterate and if you converge you will converge to λ_1, $\vec{x}^t \to \vec{v}_1$.

3) Deflate $T' \longleftarrow T - \lambda_1 \vec{v}_1 \otimes \vec{v}_1 \otimes \vec{v}_1$.

Don't know if $|\lambda_2 \vec{v}_2^T \cdot \vec{x}^0| > |\lambda_3 \vec{v}_3^T \cdot \vec{x}^0|$ but you just go on.

4) Often for the problem at hand the tensor has the underlying structure as follows

$$T = \sum_{i=1}^{K} \alpha_i \, \vec{\mu}_i \otimes \vec{\mu}_i \otimes \vec{\mu}_i$$

but the $[\vec{\mu}_1 \ldots \vec{\mu}_K]$ does <u>NOT</u> form a $\perp$ array

So in order to apply the Power Method

There is a so-called $\boxed{\text{withening process}}$

to transform T into a tensor of the form

$$\sum_{i=1}^{K} \lambda_i \, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$$

and there is a relation-ship between the

$\alpha_i \leftrightarrow \lambda_i$ & $\vec{\mu}_i \leftrightarrow \vec{v}_i$'s.

|| However knowledge of $M = \sum_{i=1}^{K} \alpha_i \vec{\mu}_i \otimes \vec{\mu}_i$ is needed.

<u>Note</u>: look for withening process / Also there is one <u>exercise</u> where you will whiten a Tensor / <u>Application</u>

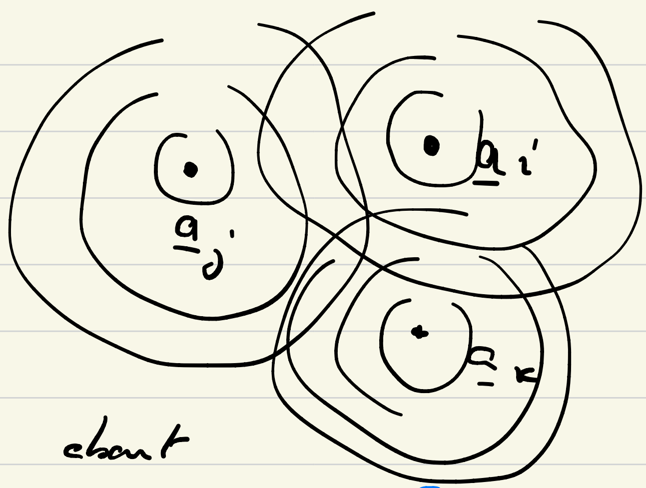
Two situations where you can apply these ideas.

1) Gaussian Mixture Models. (exercise again).

estimating a mixt of gaussians:

$$\underline{x} \in \mathbb{R}^D;\quad p(\underline{x}) = \sum_{i=1}^{K} w_i \, \frac{e^{-\frac{\|\underline{x} - \underline{a}_i\|^2}{2\sigma^2}}}{(2\pi\sigma^2)^{D/2}}.$$

$$0 \le \underline{\underline{w_i}} \le 1 \quad ; \quad \sum_{i=1}^{K} w_i = 1.$$



You showed these formulas about moment: $\sum_{i=1}^{D} \underline{e}_i \otimes \underline{e}_i$

$$\left\{\begin{aligned}
\mathbb{E}(\underline{x}) &= \sum_{i=1}^{K} w_i \, \underline{a}_i \equiv \underline{\mu} \ \ldots \\
\mathbb{E}(\underline{x} \otimes \underline{x}) &= \sigma^2 \, \boxed{I_{D\times D}} + \underbrace{\sum_{i=1}^{K} w_i \, \overbrace{\underline{a}_i \otimes \underline{a}_i}^{\underline{a}_i \underline{a}_i^T}}_{\text{if } K<D} \\
\mathbb{E}(\underline{x} \otimes \underline{x} \otimes \underline{x}) &= \sum_{i=1}^{K} w_i \, \underline{a}_i \otimes \underline{a}_i \otimes \underline{a}_i \\
&\quad + \sigma^2 \sum_{i=1}^{K} \left[\underline{\mu} \otimes \underline{e}_i \otimes \underline{e}_i + \underline{e}_i \otimes \underline{\mu} \otimes \underline{e}_i + \underline{e}_i \otimes \underline{e}_i \otimes \underline{\mu} \right]
\end{aligned}\right.$$

rank $K < D$
zero eigenvalue.

where $\underline{e}_1 \ldots \underline{e}_D$ is canonical basis of $\mathbb{R}^D$: $\underline{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ etc.

From sample of $p(\underline{x})$: $\underline{x}^{(1)} \dots \underline{x}^{(P)}$.

you can form $\underline{x}^{(j)}$; $j = 1 \dots P$.

$$M^{\alpha\beta} = \frac{1}{P} \sum_{j=1}^{P} x^{(j)\alpha} x^{(j)\beta}$$

$$\simeq \mathbb{E}(x^\alpha x^\beta) = \left(\mathbb{E}(x \otimes x)\right)^{\alpha\beta}$$

$$T^{\alpha\beta\gamma} = \frac{1}{P} \sum_{j=1}^{P} x^{(j)\alpha} x^{(j)\beta} x^{(j)\gamma}$$

$$\simeq \mathbb{E}(x^\alpha x^\beta x^\gamma) = \left(\mathbb{E}(\underline{x} \otimes \underline{x} \otimes \underline{x})\right)^{\alpha\beta\gamma}$$

__Problem:__ From $M^{\alpha\beta}$ & $T^{\alpha\beta\gamma} \longrightarrow$ deduce w_i ; $\underline{a}_i$'s.

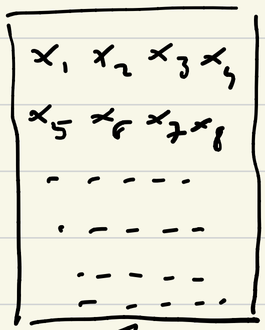
- $m^\alpha = \frac{1}{P} \sum_{j=1}^{P} x^{(j)\alpha} \longrightarrow$ you know $\underline{m}$.
- If $K < D$: σ^2 minimal eigenvalue of $M^{\alpha\beta} \longrightarrow$ you know σ^2.
- $T^{\alpha\beta\gamma}$ subtract terms with $\underline{m}$ & $\sigma^2 \rightarrow T' = \sum^{K} w_i \vec{a}_i \otimes \vec{a}_i \otimes \vec{a}_i$

Before applying Power Method / you would like to apply power method
you whiten this tensor because you here $M^{\alpha\beta}$. EXERCICE GUIDING YOU THROUGH ALL THAT

(single)

2) Topic Models for Documents.

An oversimplified model of documents (e.g. books).



(L words.)

- Collection of words x_i and we assume that $x_i \in$ Dictionary

  D possible entries.

- A document is about a single topic (sports, music, politics, cinema, confinement ...)

  $h \in \{1, \dots, K\}$ K possible topics

- Document is a random realisation of a prob distribution

$$\underline{P(x_1 \dots x_L, h)} \equiv \underbrace{p(h)}\prod_{i=1}^{L}\underbrace{p(x_i \mid h)}.$$

Given a topic the words are all independent.

You have access to documents where you get to see words only (don't know the topic): You get are empirical frequencies for

$P(x_1)$; $P(x_1, x_2)$; $P(x_1, x_2, x_3)$ ect ...

Translate this learning problem in the language of tensors:

x = word $\in$ Dictionary of dimension D.

$$= \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_D\} \text{ words.}$$

• encoding of words as canonical basis vectors

$$\text{line } \alpha \longrightarrow \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \vec{e}_\alpha \in \mathbb{R}^D.$$

• $P(\underline{x} = \underline{e}_\alpha \mid h = i) \equiv \mu_i^\alpha$

cond prob of word $\underline{e}_\alpha$ given topic $i \in \{1 \dots K\}$.

$$\underline{\mu}_i = \begin{pmatrix} \mu_i^1 \\ \mu_i^2 \\ \vdots \\ \mu_i^D \end{pmatrix} = \text{vector of cond probabilities of words given Topic } i.$$

• $P(h = i) \equiv w_i$

$$\boxed{\begin{array}{l} \text{Learning Task: estimate } w_i \text{ \& } \underline{\mu}_i \\ \text{from } \underbrace{P(x_1)\,;\; P(x_1, x_2)\,;\; P(x_1, x_2, x_3)}_{\text{empirical frequencies}} \end{array}}$$

This now can be put in the form of determining a tensor decomposition: Look at Moment:

$$\mathbb{E}(\underline{x}_1) = \sum_{\alpha=1}^{D} \underline{e}_\alpha \, p(\underline{x}_1 = \underline{e}_\alpha)$$

$$= \sum_{\alpha=1}^{D} \underline{e}_\alpha \sum_{i=1} p(\underline{x}_1 = e_\alpha \,|\, i)\, w_i$$

$$= \sum_{i=1}^{K} w_i \left\{ \sum_{\alpha=1}^{D} \underbrace{p(x_1 = e_\alpha \,|\, i)}_{\mu_i^\alpha} \, \underbrace{\underline{e}_\alpha}_{\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix} \leftarrow \alpha} \right\}$$

$$\{\cdots\} = \underline{\mu}_i$$

$$\Rightarrow \boxed{\mathbb{E}(\underline{x}_1) = \sum_{i=1}^{K} w_i \, \underline{\mu}_i}$$

in practice you have an empirical estimate for $\mathbb{E}(\underline{x}_1)$.

$$\mathbb{E}(\underline{x}_1 \otimes \underline{x}_2) = \sum_{\alpha=1}^{D} \sum_{\beta=1}^{D} \underline{e}_\alpha \otimes \underline{e}_\beta \, P(\underline{x}_1 = \underline{e}_\alpha ; \underline{x}_2 = \underline{e}_\beta)$$

$$= \sum_{\alpha,\beta=1}^{D} \underline{e}_\alpha \otimes \underline{e}_\beta \sum_{i=1}^{K} w_i \, p(x_1 = e_\alpha, x_2 = e_\beta \mid i)$$

$$= \sum_{i=1}^{K} w_i \sum_{\alpha,\beta=1}^{D} \underbrace{p(x_1 = e_\alpha, x_2 = e_\beta \mid i)} \, \underline{e}_\alpha \otimes \underline{e}_\beta .$$

Recall the fresh Model for document:

$$\underbrace{p(x_1 = \alpha \mid i)}_{\mu_i^\alpha} . \underbrace{p(x_2 = \beta \mid i)}_{\mu_i^\beta} .$$

$$= \sum_{i=1}^{K} w_i \underbrace{\sum_{\alpha,\beta=1}^{D} \mu_i^\alpha \underline{e}_\alpha \otimes \mu_i^\beta \underline{e}_\beta}.$$

$$= \sum_{i=1}^{K} w_i \sum_{\alpha=1}^{D} \mu_i^\alpha \underline{e}_\alpha \otimes \sum_{\beta=1}^{D} \mu_i^\beta \underline{e}_\beta .$$

$$\boxed{\mathbb{E}(\underline{x}_1 \otimes \underline{x}_2) = \sum_{i=1}^{K} w_i \, \underline{\mu}_i \otimes \underline{\mu}_i}$$

$\mathbb{E}(x_1 \otimes x_2)$ is accessible for empirical obs.

Finally same calculation:

$$\boxed{\mathbb{E}(\underline{x}_1 \otimes \underline{x}_2 \otimes \underline{x}_3) = \sum_{i=1}^{K} w_i \, \underline{\mu}_i \otimes \underline{\mu}_i \otimes \mu_i}$$

again this is accessible from empirical observation of documents.

In summary you have

$$\begin{cases} \mathbb{E}(x \otimes \underline{x}) = \sum_{i=1}^{K} w_i \, \underline{\mu}_i \otimes \underline{\mu}_i & \text{Matrix.} \\ \mathbb{E}(\underline{x} \otimes \underline{x} \otimes \underline{x}) = \sum_{i=1}^{K} w_i \, \underline{\mu}_i \otimes \underline{\mu}_i \otimes \mu_i \end{cases}$$

- Apply whitening process to turn the 3rd order Tensor into $\sum \lambda_i \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$ with $\vec{v}_i$'s $\perp$.
- Apply Tensor Power Method to get λ_i's, $\vec{v}_i$'s
- Recover w_i & $\underline{\mu}_i = (\mu_i^\alpha)$

  $w_i = \mathbb{P}(h = i)$, $\mu_i^\alpha = \mathbb{P}(e_\alpha \mid i)$.