TENSOR METHODS IN LEARNING THEORY

Time line of these lectures:

1) Motivations & typical probl → tensors appear
2) Introduce a language & basic def about tensors.
3) Tensor Rank / Notion of Tensor Decomposition

   Tool to analyze data that has a "tensor structure"
4) Thms & Algos on tensor decomposition

   Jennrich's Thm ; Alt Least Square Algo ;

   tensor power method.
5) Higher order SVD for tensors
6) Applications to problems: Gaussian Mixture Models ;

   Topic Models ; Multiview classification.

1. Motivates & 2 typical problems.

a) Estimate parameters of Gaussian Mixture Model.

$\underline{x} \in \mathbb{R}^D$ $\quad P(\underline{x}) = \sum_{i=1}^{K} w_i \underbrace{\exp\left\{-\frac{\|\underline{x} - \underline{a}_i\|^2}{2\underline{\sigma}^2}\right\}}_{(2\pi\sigma)^{-D/2}}$.

$$\begin{cases} 0 \leq w_i \leq 1 \\ \sum_{i=1}^{K} w_i = 1. \end{cases}$$

Problem following: Have access samples $\underline{x}^{(1)} \dots \underline{x}^{(N)}$

from there → Learn param of model:
$\{ w_i , \underline{a}_i , \underline{\underline{\sigma}} \}$

Popular idea: try to match moments.

second moment (exercise)

$$M_2^{th} = \mathbb{E}(\underline{x}\,\underline{x}^T) = \sigma^2 I_{D\times D} + \sum_{i=1}^{K} w_i \overbrace{\underline{a}_i \underline{a}_i^T}^{\text{rank-one matrices}}$$

Match this to empirical moment

$$\frac{1}{N} \sum_{m=1}^{N} \underline{x}^{(m)} \underline{x}^{(m)T} = M_2^{emp.}$$

$$\mathbb{E}(\underline{x}\,\underline{x}^T) = \sigma^2 I_{D\times D} + \sum_{i=1}^{K} w_i \underbrace{\underline{a}_i \underline{a}_i^T}_{\text{rank one}}$$

pos semi def.

pos semi-def $\overbrace{\text{matrix}}^{D\times D}$ with rank at most K.

Suppose in usual applications $K < D$.

$\exists$ zero eigenvalues which are the min e.v.
Thus $\mathbb{E}(\underline{x}\,\underline{x}^T)$ has σ^2 as its min e.v.

First thing: Estimate σ^2 by the min e.v of $M_2^{emp} = \frac{1}{N}\sum_{n=1}^{N} \underline{x}^{(n)} \underline{x}^{(n)T}$

Second thing: Estimate w_i & $\underline{a}_i$ from the matrix $M_2^{emp} - \sigma^2 I_{D\times D} \approx \sum_{i=1}^{K} w_i \underline{a}_i \underline{a}_i^T$

you have to DECOMPOSE A MATRIX IN A SUM OF RANK ONE MATRICES

also called "Matrix decomposition or factorisation problem".

You run into a problem here because such

decomp are usually typ not unique.

To see non unicity take first $w_i = \frac{1}{K}$

set $\frac{1}{\sqrt{K}} \underline{a}_i = \underline{b}_i$. Decompose

$$\sum_{i=1}^{K} \underline{b}_i \underline{b}_i^T = \overbrace{\underset{D \times K}{\left[\underline{b}_1 \ldots \underline{b}_K \right]}}^{B} \underset{K \times D}{\begin{bmatrix} \underline{b}_1^T \\ \vdots \\ \underline{b}_K^T \end{bmatrix}}.$$

$$= B\, B^T$$

such factorization is Not unique

$$= B \underbrace{R R^T}\, B^T$$

where R is a $K \times K$ rotation matrix i.e $R R^T = I_{D \times D}$.

$$= B' B'^T \quad \text{with} \quad B' = B R.$$

the "rotation problem" makes such fact non unique.

- How would you somehow determine the true solution $\{w_i, \underline{a}_i\}$ $i = 1 \dots K$?

Match Moments also for higher Moments

exercise to show

$$M_3^{th} = \mathbb{E}\left(x^\alpha x^\beta x^\gamma\right) = \sum_{i=1}^{K} w_i \, a_i^\alpha \, a_i^\beta \, a_i^\gamma$$

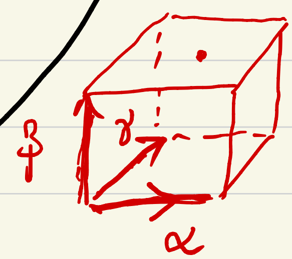
$$\underline{x} = (x^1, x^2, \dots, x^D) \qquad \underline{a}_i = (a_i^1, \dots, a_i^D)$$

$$\alpha, \beta, \gamma \in \{1 \dots D\}$$

$$M_3^{emp} = \frac{1}{N} \sum_{m=1}^{N} x^{(m)\alpha} \, x^{(m)\beta} \, x^{(m)\gamma}$$

# of sample $m = 1 \dots N$

$$\underline{x}^{(m)} = (x^{(m)1}, \dots, x^{(m)D}).$$



Take $(M_3^{emp})^{\alpha\beta\gamma}$ try to factor it as

$$\sum_{i=1}^{K} w_i \, a_i^\alpha \, a_i^\beta \, a_i^\gamma .$$

Set of D^3 numbers is called a Tensor.

TENDS TO HAVE UNIQUE SOLUTION

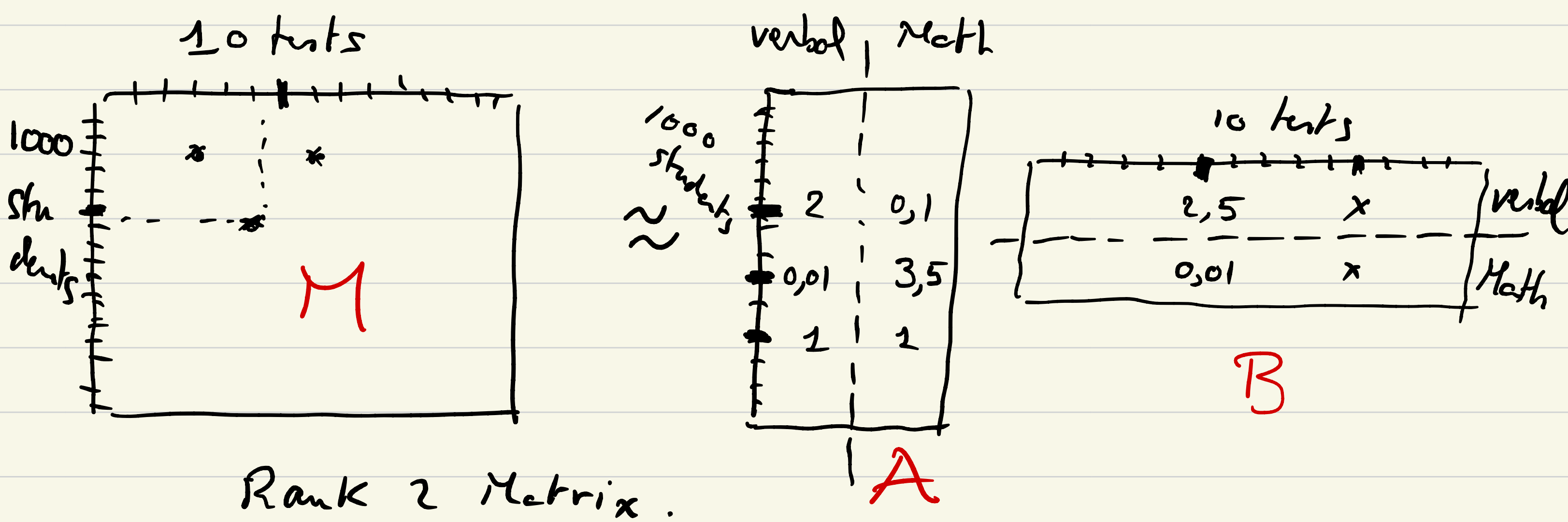
b) Second Motivating Problem: Spearman's probl.

(psychologist 20-th century.)

devise ways to "measure" "intelligence".

hyp: two kind of intelligence $\to$ verbal, $\to$ mathematical.

data available: comes as a Matrix



Rank 2 Matrix.

|| You extract verbal/math abilities through the factorization of the Matrix.

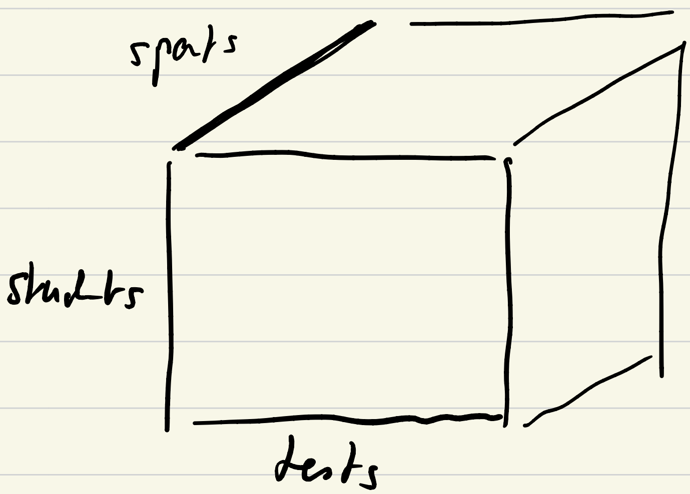
Again you have the "Rotation Problem"

$$\begin{cases} M = AB = \underbrace{AR}_{A'}\underbrace{R^TB}_{B'} \quad \ldots \\ \text{Hard to interpret verbal/math abilities.} \end{cases}$$

$R = 2 \times 2$ orth matrix $RR^T = 1$

Possible solution: you may have to take into account a third ability (e.g. ability to socialize or make sports ....)

→ introduce a third dimension in the data.



try to factor this tensor somehow according to three directions.

and this typically has a unique solution

#.

Enough motivation to talk about tensors.

↙

multidimensional arrays of numbers

Remark: Matrix fact or decomposition has a non-unique solution. (rotation problem).

BUT often under additional constraints you have a unique solution.

e.g imagine M is a sym rank-one matrix

$\Rightarrow M = \underline{a}\,\underline{a}^T$, $\underline{a}$ is unique up to a sign.

e.g any real $M = U \Sigma V^T$

$\uparrow$

SVD

here U & V are orthogonal and Σ diag with singular value.

If all sing values $\sigma_1 > \sigma_2 \dots > \sigma_R > 0$ are distinct & non zero then the decomp is unique:

$$M = \sum_{i=1}^{R} \sigma_i \underline{u}_i \underline{v}_i^T$$

constraint orthogonality of set of vectors.

$U = [\underline{u}_1 \dots \underline{u}_R]$

$V = [\underline{v}_1 \dots \underline{v}_R]$.

2) $\overbrace{\text{Introduction}}^{\text{first}}$ to the language of Tensors.

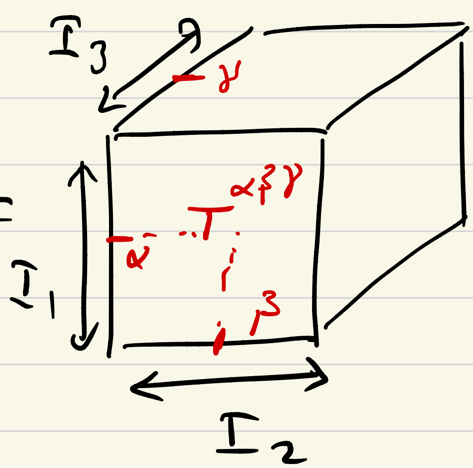
- Concrete approach using arrays of numbers.

* A column vect $\underline{x} = \begin{bmatrix} x_{\underline{1}} \\ \vdots \\ x_{\underline{D}} \end{bmatrix} \in \mathbb{R}^D$: one index

* A matrix $X = \begin{array}{c|c} & \beta \; (I_2) \\ \hline \alpha \; (I_1) & \begin{bmatrix} X_{\alpha\beta} \end{bmatrix} \end{array} \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2}$

two index object

* A $\underbrace{\text{tensor}}_{}^{\text{of order 3.}}$ is a multidimensional array of numbers (later on more math def).

$T =$ 

$T^{\alpha\beta\gamma}$ has 3 indices.
3-index object.

$\alpha = 1 \ldots I_1$
$\beta = 1 \ldots I_2$
$\gamma = 1 \ldots I_3$

$\in$
$\mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}$.

* Order-p tensor is just a p-dimensional array of numbers

$$T^{\alpha_1 \alpha_2 \dots \alpha_p} \in \mathbb{R}^{I_1} \times \dots \times \mathbb{R}^{I_p} . \quad \begin{cases} p = 3 & \text{cube} \\ p = 4 \dots & \text{hypercube.} \end{cases}$$

$\alpha_1 \in \{1 \dots I_1\}$
$\alpha_2 \in \{1 \dots I_2\}$
$\vdots$
$\alpha_p \in \{1 \dots I_p\}$

Terminology. (varies in litterature)

- p is called the "order", the way the number of "modes",
- $p = 1 \to$ vector
  $p = 2 \to$ matrix (2-tensor).
  $p \geq 3 \to$ tensor.

PAUSE