Homework 10 (Graded Homework 4): due Friday, June 3 2022
CS-526 Learning Theory

Note: The tensor product is denoted by $\otimes$. In other words, for vectors $\vec{a}, \vec{b}, \vec{c}$ we have that $\vec{a} \otimes \vec{b}$ is the square array $a^\alpha b^\beta$ where the superscript denotes the components, and $\vec{a} \otimes \vec{b} \otimes \vec{c}$ is the cubic array $a^\alpha b^\beta c^\gamma$. We denote components by superscripts because we need the lower index to label vectors themselves.

Problem 1: Comparison of tensor rank and multilinear rank

Recall that the "tensor rank" (usually called "rank") is the smallest R such that the multi-array $T^{\alpha\beta\gamma}$ can be decomposed as a sum of rank one terms in the form

$$T^{\alpha\beta\gamma} = \sum_{j=1}^{R} a_j^\alpha b_j^\beta c_j^\gamma \qquad \text{or equivalently} \qquad T = \sum_{j=1}^{R} \vec{a}_j \otimes \vec{b}_j \otimes \vec{c}_j \ .$$

This is often denoted $\mathrm{rank}_\otimes(T) = R$. On the other hand, the multilinear rank is the tuple $\mathrm{rank}_\boxplus(T) = (R_1, R_2, R_3)$ where R_1, R_2, R_3 are the ranks of the three matricizations $T_{(1)}$, $T_{(2)}$, $T_{(3)}$ defined in class.

1. Show that $\max \mathrm{rank}_\boxplus(T) \leq \mathrm{rank}_\otimes(T)$.

Problem 2: Non-unicity of the Tucker decomposition

Let $T = (T^{\alpha\beta\gamma})$, $\alpha = 1, \dots I_1$, $\beta = 1, \dots I_2$, $\gamma = 1, \dots I_3$ an order-three tensor. Suppose that its multilinear rank is $\mathrm{rank}_\boxplus(T) = (R_1, R_2, R_3)$ which means that R_1, R_2, R_3 are the ranks of the three matricizations $T_{(1)}$, $T_{(2)}$, $T_{(3)}$ defined in class. We have seen in class that any such tensor has a so-called *Tucker decomposition* (also called higher order singular value decomposition):

$$T = \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr}\, \vec{u}_p \otimes \vec{v}_q \otimes \vec{w}_r$$

where each of the matrices $[\vec{u}_1, \cdots, \vec{u}_{R_1}]$, $[\vec{v}_1, \cdots, \vec{v}_{R_2}]$, $[\vec{w}_1, \cdots, \vec{w}_{R_3}]$ are made of orthogonal unit vectors. $G = (G^{pqr})$ is called the core tensor (and is not diagonal in general). In this problem you will prove that this decomposition is not unique and, in fact, that there exist an infinity of such decompositions related by orthogonal transformations.

Let $M^{(u)} = (M^{(u)}_{pp'})$, $M^{(v)} = (M^{(v)}_{qq'})$ and $M^{(w)} = (M^{(w)}_{rr'})$ be three orthogonal matrices of dimensions $R_1 \times R_1$, $R_2 \times R_2$ and $R_3 \times R_3$. Define the vectors:

$$\vec{x}_{p'} = \sum_{p=1}^{R_1} M^{(u)}_{p'p} \vec{u}_p, \quad \vec{y}_{q'} = \sum_{p=1}^{R_2} M^{(v)}_{q'q} \vec{v}_q, \quad \vec{z}_{r'} = \sum_{p=1}^{R_3} M^{(w)}_{r'r} \vec{w}_r \ .$$

Show that there exist a core tensor $H = (H^{pqr})$ of dimension $R_1 \times R_2 \times R_3$ such that

$$T = \sum_{p,q,r=1}^{R_1,R_2,R_3} H^{pqr}\, \vec{x}_p \otimes \vec{y}_q \otimes \vec{z}_r \ .$$

Problem 3: Tensor decomposition & estimation of a sensing matrix

Let $N \geq K$ two positive integers. Define the $N \times K$ real matrix $A := \begin{bmatrix} \vec{\mu}^{(1)} & \vec{\mu}^{(2)} & \cdots & \vec{\mu}^{(K)} \end{bmatrix}$ where the column vectors $\vec{\mu}^{(k)} = (\mu^k_\alpha)_{\alpha=1}^N$, $k = 1 \ldots K$, are (fixed) N-dimensional linearly independent vectors.

Let $\vec{h} = (h_k)_{k=1}^K$ be a random vector whose components h_k's are independently (but not necessarily identically) distributed. We assume that $\forall k : \mathbb{E}[h_k] = \mathbb{E}[h_k^3] = 0$ and $\mathbb{E}[h_k^2], \mathbb{E}[h_k^4]$ are finite positive. We define the *excess kurtoses* $\mathcal{K}_k = \frac{\mathbb{E}[h_k^4]}{\mathbb{E}[h_k^2]^2} - 3$. If h_k has a zero-mean Gaussian distribution then $\mathcal{K}_k = 0$, so $\mathcal{K}_k$ can be essentially viewed as a measure of non-Gaussianity.

We are given L observations $\vec{y}^{(\ell)} = (y^\ell_\alpha)_{\alpha=1}^N := A\vec{h}^{(\ell)}$ where $\vec{h}^{(1)}, \vec{h}^{(2)}, \ldots, \vec{h}^{(L)} \overset{\text{i.i.d.}}{\sim} \vec{h}$. Except for what is known on the distribution of $\vec{h}$, we don't know anything on the input vectors $\vec{h}^{(1)}, \vec{h}^{(2)}, \ldots, \vec{h}^{(L)}$. The goal of the exercise is to show how to recover the columns of the sensing matrix A from these L observations.

1. Let $\vec{y} := A\vec{h}$. We define $\widehat{S}$ and $\widehat{F}$ the empirical estimates (using the L observations $\vec{y}^{(\ell)}$) of the second-moment matrix $S := \mathbb{E}[\vec{y} \otimes \vec{y}]$ and the fourth-moment tensor $F := \mathbb{E}[\vec{y} \otimes \vec{y} \otimes \vec{y} \otimes \vec{y}]$.

   Write down expressions for the components $\widehat{S}_{\alpha\beta}$ of $\widehat{S}$ and $\widehat{F}_{\alpha\beta\gamma\delta}$ of $\widehat{F}$ in terms of the components of $\vec{y}^{(1)}, \vec{y}^{(2)}, \ldots, \vec{y}^{(L)}$.

2. From now on we suppose that $\widehat{S}$ and $\widehat{F}$ are good estimates of S and F, respectively. Prove the following identities:

$$S = \sum_{k=1}^K \mathbb{E}[h_k^2]\, \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \ ;$$

$$F = \sum_{k=1}^K \mathbb{E}[h_k^4]\, \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} + \sum_{1 \leq j \neq k \leq K} \mathbb{E}[h_j^2]\mathbb{E}[h_k^2] \Big(\vec{\mu}^{(j)} \otimes \vec{\mu}^{(j)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} + \vec{\mu}^{(j)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(j)} \otimes \vec{\mu}^{(k)} + \vec{\mu}^{(j)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(j)} \Big).$$

3. We now form the tensor T with components

$$T_{\alpha\beta\gamma\delta} := F_{\alpha\beta\gamma\delta} - S_{\alpha\beta}S_{\gamma\delta} - S_{\alpha\gamma}S_{\beta\delta} - S_{\alpha\delta}S_{\beta\gamma} \ .$$

   Use the previous question to show that

$$T = \sum_{k=1}^K \mathcal{K}_k \mathbb{E}[h_k^2]^2\, \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \otimes \vec{\mu}^{(k)} \ .$$

4. Show that $S = UDU^T$ with $U = \begin{bmatrix} \vec{u}^{(1)} & \vec{u}^{(2)} & \cdots & \vec{u}^{(K)} \end{bmatrix} \in \mathbb{R}^{N\times K}$ a matrix with orthonormal columns, and $D = \mathrm{Diag}(d_1, d_2, \ldots, d_K)$ a diagonal matrix with diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_K > 0$.

5. Define the vectors $\vec{v}^{(k)} := \sqrt{\mathbb{E}[h_k^2]}\, W^T \vec{\mu}^{(k)}$, $k = 1 \ldots K$, where $W = UD^{-\frac{1}{2}}$ and the tensor $\widetilde{T} := \sum_{k=1}^{K} \mathcal{K}_k\, \vec{v}^{(k)} \otimes \vec{v}^{(k)} \otimes \vec{v}^{(k)} \otimes \vec{v}^{(k)}$.

   Explain how to obtain the components $\widetilde{T}_{\alpha\beta\gamma\delta}$ of $\widetilde{T}$ from those of T, i.e., write down the formula relating them. How is this process (the transformation of T into $\widetilde{T}$) called?

6. As we have seen in class, the set of vectors $\{\vec{v}^{(1)}, \ldots, \vec{v}^{(K)}\}$ is orthonormal and we can try to recover them using the tensor power method.

   What happens if one of the excess kurtosis $\mathcal{K}_k$ is zero?

7. From now on we suppose that all the excess kurtoses are nonzero.

   Write a small pseudo-code for the power method applied to $\widetilde{T}$ to recover $\mathcal{K}_k$ and (up to a plus or minus sign) $\vec{v}^{(k)}$ for $k = 1 \ldots K$.

8. Assume that we also know the second moments $\mathbb{E}[h_k^2]$ for $k = 1 \ldots K$.

   After having recovered $\mathcal{K}_k$ and $\pm\vec{v}^{(k)}$ for $k = 1 \ldots K$ with the power method, how do you recover $\pm\vec{\mu}^{(k)}$ (so up to a plus or minus sign) for $k = 1 \ldots K$?