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

$$T^{\alpha \beta \gamma} = \sum_{j=1}^{R} a^\alpha_j b^\beta_j c^\gamma_j \quad \text{or equivalently} \quad T = \sum_{j=1}^{R} \vec{a}_j \otimes \vec{b}_j \otimes \vec{c}_j.$$

This is often denoted $\text{rank}_{\otimes}(T) = R$. On the other hand, the multilinear rank is the tuple $\text{rank}_{\boxtimes}(T) = (R_1, R_2, R_3)$ 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 $\max \text{rank}_{\boxtimes}(T) \leq \text{rank}_{\otimes}(T)$.

Problem 2: Non-unicity of the Tucker decomposition

Let $T = (T^{\alpha \beta \gamma})$, $\alpha = 1, \ldots, I_1$, $\beta = 1, \ldots, I_2$, $\gamma = 1, \ldots, I_3$ an order-three tensor. Suppose that its multilinear rank is $\text{rank}_{\boxtimes}(T) = (R_1, R_2, R_3)$ 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 $[\vec{u}_1, \ldots, \vec{u}_{R_1}]$, $[\vec{v}_1, \ldots, \vec{v}_{R_2}]$, $[\vec{w}_1, \ldots, \vec{w}_{R_3}]$ 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:

$$\vec{x}_{p'} = \sum_{p=1}^{R_1} M^{(u)}_{p'p} \vec{u}_p, \quad \vec{y}_{q'} = \sum_{q=1}^{R_2} M^{(v)}_{q'q} \vec{v}_q, \quad \vec{z}_{r'} = \sum_{r=1}^{R_3} M^{(w)}_{r'r} \vec{w}_r.$$
Show that there exist a core tensor $H = (H^{pqr})$ of dimension $R_1 \times R_2 \times R_3$ such that

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

**Problem 3: Whitening of a tensor**

Consider the tensor

$$T = \sum_{i=1}^{K} \lambda_i \bar{\mu}_i \otimes \bar{\mu}_i \otimes \bar{\mu}_i$$

where $\bar{\mu}_i \in \mathbb{R}^D$ are linearly independent (so $K \leq D$) and $\lambda_i$ are strictly positive. Consider the matrix

$$M = \sum_{i=1}^{K} \lambda_i \bar{\mu}_i \otimes \bar{\mu}_i = \sum_{i=1}^{K} \lambda_i \bar{\mu}_i \bar{\mu}_i^T.$$ 

Note that this is a rank-$K$ symmetric positive semi-definite matrix (there are $D - K$ zero eigenvalues). Denote $d_1 \geq d_2 \geq \cdots \geq d_K$ the strictly positive eigenvalues of $M$ and $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_K$ the corresponding eigenvectors. Hence $M = UD\text{diag}(d_1, \ldots, d_K)U^T$ where $U = [\bar{u}_1 \ \bar{u}_2 \ \cdots \ \bar{u}_K]$. Define the $D \times K$ matrix:

$$W = UD\text{diag}(d_1^{-1/2}, d_2^{-1/2}, \ldots, d_K^{-1/2}).$$

The whitening of $T$ is defined as the new tensor obtained by the multilinear transform

$$T(W, W, W) := \sum_{i=1}^{K} \nu_i \bar{\nu}_i \otimes \bar{\nu}_i \otimes \bar{\nu}_i$$

where $\nu_i = \lambda_i^{-1/2}$ and $\bar{\nu}_i = \sqrt{\lambda_i} W^T \bar{\mu}_i$.

1. Show that $W^T MW = I$ where $I$ is the $K \times K$ identity matrix. Deduce that the $\bar{\nu}_i$’s are orthonormal, i.e., $V^T V = I$ where $V = [\bar{\nu}_1 \ \bar{\nu}_2 \ \cdots \ \bar{\nu}_K]$.

2. Suppose we are given a tensor $T$ of the form $T = \sum_{i=1}^{K} \lambda_i \bar{\mu}_i \otimes \bar{\mu}_i \otimes \bar{\mu}_i$ and a matrix $M = \sum_{i=1}^{K} \lambda_i \bar{\mu}_i \bar{\mu}_i^T$ where $\bar{\mu}_i \in \mathbb{R}^D$ are linearly independent and $\lambda_i > 0$. Explain how applying the tensor power method to the whitened tensor $T(W, W, W)$ helps you recover the $\lambda_i$’s and $\mu_i$’s, and give a closed-form formula for the matrix $\mu = [\bar{\mu}_1 \ \cdots \ \bar{\mu}_K]$ that uses $V$, $\text{Diag}(\nu_1, \ldots, \nu_K)$ and $W$.

**Problem 4: Estimating parameters of a Gaussian Mixture Model**

Consider the following Gaussian Mixture Model (GMM):

$$p(x) = \sum_{i=1}^{K} w_i \frac{1}{(2\pi\sigma^2)^{D/2}} \exp \left( -\frac{\|x - a_i\|^2}{2\sigma^2} \right)$$

where $x \in \mathbb{R}^D$, the matrix $A := [a_1 \ \cdots \ a_K]$ has full column rank, i.e., its columns are linearly independent, and the weights $w_i \in (0, 1]$ satisfy $\sum_{i=1}^{K} w_i = 1$. We suppose that
\( K < D \). You can think of the distribution \( p(\cdot) \) has \( K \) clusters centered at \( a_1, \ldots, a_K \), and the probability to belong to the cluster centered at \( a_r \) is \( w_r \).

You saw in Problem 1 of Homework 6 that the first moment vector, the second moment matrix and the third moment tensor of this distribution are:

\[
\mathbb{E}[x] = \sum_{i=1}^{K} w_i a_i
\]

\[
\mathbb{E}[xx^T] = \sigma^2 I_D + \sum_{i=1}^{K} w_i a_i a_i^T
\]

\[
\mathbb{E}[xx \otimes xx] = \sum_{i=1}^{K} w_i a_i \otimes a_i \otimes a_i + \sigma^2 \sum_{j=1}^{D} \mathbb{E}[x] \otimes e_j \otimes e_j + e_j \otimes e_j \otimes \mathbb{E}[x]
\]

You will find on the course webpage the notebook `problem_4_notebook.ipynb` and a NumPy binary file `data.npz` with datapoints drawn from a GMM. This notebook will guide through a method to recover the matrix \( A \) and the weights \( w_i \) of the model in order to perform classification.

We ask you to fill this notebook where needed (check for the “#TO DO”!). When submitting your homework you should also include the filled notebook that you will rename `problem_4_FIRSTNAME_LASTNAME.ipynb`. When grading your homework we will run your entire notebook with `Kernel → Restart & Run All` in the notebook menu bar. If you don’t have Python on your computer, you can access the notebook on Google Colab: https://colab.research.google.com/drive/1rK_82P4llmEhqUf0AzDl2CqnxRinyHHT

Please answer the question below:

1. Show that the minimum eigenvalue of \( \mathbb{E}[xx^T] \) is \( \sigma^2 \).