Homework 11: 1 June 2021
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: Whitening of a tensor

Consider the tensor

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

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

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

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 $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_K$ the corresponding eigenvectors. Hence $M = U \mathrm{Diag}(d_1, \ldots, d_K) U^T$ where $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_K \end{bmatrix}$. Define the $D \times K$ matrix:

$$W = U \mathrm{Diag}(d_1^{-1/2}, d_2^{-1/2}, \cdots, d_K^{-1/2}) \ .$$

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

$$T(W, W, W) := \sum_{i=1}^{K} \lambda_i \, (W^T \vec{\mu}_i) \otimes (W^T \vec{\mu}_i) \otimes (W^T \vec{\mu}_i) = \sum_{i=1}^{K} \nu_i \, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$$

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

1. Show that $W^T M W = I$ where I is the $K \times K$ identity matrix. Deduce that the $\vec{v}_i$'s are orthonormal, i.e., $V^T V = I$ where $V = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_K \end{bmatrix}$.

2. Suppose we are given a tensor T of the form $T = \sum_{i=1}^{K} \lambda_i \, \vec{\mu}_i \otimes \vec{\mu}_i \otimes \vec{\mu}_i$ and a matrix $M = \sum_{i=1}^{K} \lambda_i \, \vec{\mu}_i \vec{\mu}_i^T$ where $\vec{\mu}_i \in \mathbb{R}^D$ 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 λ_i's and μ_i's, and give a closed-form formula for the matrix $\mu = \begin{bmatrix} \vec{\mu}_1 & \cdots & \vec{\mu}_K \end{bmatrix}$ that uses V, $\mathrm{Diag}(\nu_1, \ldots, \nu_K)$ and W.