Homework 8: 21 April 2020
CS-526 Learning Theory

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

Problem 1: A multiple choice question

Find the correct answer(s).

Let $w_i(\epsilon)$ for $i \in \{1, \dots, K\}$ be continuous functions of $\epsilon \in [0,1]$. Suppose that for all $\epsilon \in [0,1]$ the $N \times K$ matrices $\begin{bmatrix}\mathbf{a}_1 + \epsilon \mathbf{a}'_1 & \cdots & \mathbf{a}_K + \epsilon\, \mathbf{a}'_K\end{bmatrix}$, $\begin{bmatrix}\mathbf{b}_1 + \epsilon\, \mathbf{b}'_1 & \cdots & \mathbf{b}_K + \epsilon\, \mathbf{b}'_K\end{bmatrix}$ and $\begin{bmatrix}\mathbf{c}_1 + \epsilon\, \mathbf{c}'_1 & \cdots & \mathbf{c}_K + \epsilon\, \mathbf{c}'_K\end{bmatrix}$ have rank K. Consider the tensor

$$T(\epsilon) = \sum_{i=1}^{K} w_i(\epsilon)\, (\mathbf{a}_i + \epsilon \mathbf{a}'_1) \otimes (\mathbf{b}_i + \epsilon \mathbf{b}'_1) \otimes (\mathbf{c}_i + \epsilon \mathbf{c}'_1)\ .$$

(A) The tensor rank equals K for all $\epsilon \in [0,1]$.

(B) The tensor rank equals K for all $\epsilon \in [0,1]$ such that $\forall i \in \{1, \dots, K\} : w_i(\epsilon) \neq 0$.

(C) It may happen that the tensor rank of the limit $\lim_{\epsilon \to 0} T(\epsilon)$ is $K+1$.

(D) If we replace the assumption that $\begin{bmatrix}\mathbf{c}_1 + \epsilon\, \mathbf{c}'_1 & \cdots & \mathbf{c}_K + \epsilon\, \mathbf{c}'_K\end{bmatrix}$ is rank K by the assumption that these vectors are pairwise independent, then the tensor rank can *never* be K whatever the assumptions on $w_i(\epsilon)$, $i = 1, \dots, K$.

Problem 2: A simultaneous diagonalization method for tensor decomposition

Let $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ a set of k *linearly independent* column vectors of dimension n (with real components). We will assume throughout the problem that these vectors have *unit norms.* Set

$$T_2 = \sum_{i=1}^{k} w_i\, \mathbf{a}_i \otimes \mathbf{a}_i\ , \quad T_3 = \sum_{i=1}^{k} w_i\, \mathbf{a}_i \otimes \mathbf{a}_i \otimes \mathbf{a}_i$$

where w_i, $i = 1, \dots, k$, are nonzero real numbers.

We are given the arrays of components $T_2^{\alpha\beta}$, $T_3^{\alpha\beta\gamma}$, $\alpha, \beta, \gamma \in \{1, \dots, n\}$ and want to determine $w_1, \cdots, w_k$ as well as $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$. This problem guides you through a method that uses the simultaneous diagonalization of appropriate matrices to do so.

The following multilinear transformation of T_3 will be used:

$$T_3(I, I, \mathbf{u}) = \sum_{i=1}^{k} w_i (\mathbf{a}_i \otimes \mathbf{a}_i)(\mathbf{u}^T \mathbf{a}_i)\ ,$$

where I denotes the identity matrix and $\mathbf{u}$ is an n-dimensional real column vector ($\mathbf{u}^T$ is its transpose).

1. Define the $n \times k$ matrix $V = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_k \end{bmatrix}$. Show that

$$T_2 = V\,\mathrm{Diag}(w_1, \ldots, w_k)V^T$$
$$T_3(I, I, \mathbf{u}) = V\,\mathrm{Diag}(w_1, \ldots, w_k)\mathrm{Diag}(\mathbf{u}^T\mathbf{a}_1, \ldots, \mathbf{u}^T\mathbf{a}_k)V^T$$

   where $\mathrm{Diag}(x_1, \ldots, x_k)$ is the diagonal matrix with x_i's on the diagonal.

2. Now we specialize to the case $n = k$. Why is T_2 an invertible matrix?

3. We choose $\mathbf{u}$ from a continuous distribution over $\mathbb{R}^n$. Still in the case $n = k$.

   a) Explain how you can almost surely recover the set of $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ (up to a plus or minus sign in front of the $\mathbf{a}_i$'s) from the matrix

$$M = T_3(I, I, \mathbf{u})T_2^{-1}$$

   using standard linear algebra methods.

   b) How do you then recover the w_i's?

Problem 3: Kronecker, Khatri-Rao, Hadamard products: check useful identities

We recall a few definitions seen in class. The Kronecker product of two column vectors $\mathbf{b} \in \mathbb{R}^I$ and $\mathbf{c} \in \mathbb{R}^J$ is the column vector:

$$\mathbf{c} \otimes_{\mathrm{Kro}} \mathbf{b} \triangleq \begin{bmatrix} c_1\mathbf{b}^T & c_2\mathbf{b}^T & \cdots & c_J\mathbf{b}^T \end{bmatrix}^T .$$

The Kronecker product of two row vectors $\mathbf{d}$ and $\mathbf{e}$ is the row vector:

$$\mathbf{d} \otimes_{\mathrm{Kro}} \mathbf{e} \triangleq \begin{bmatrix} d_1\mathbf{e} & d_2\mathbf{e} & \cdots & d_J\mathbf{e} \end{bmatrix} .$$

The Khatri-Rao product of two matrices $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_R \end{bmatrix} \in \mathbb{R}^{I \times R}$ and $C = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_R \end{bmatrix} \in \mathbb{R}^{J \times R}$ is the $(IJ) \times R$ matrix:

$$C \otimes_{\mathrm{Khr}} B \triangleq \begin{bmatrix} \mathbf{c}_1 \otimes_{\mathrm{Kro}} \mathbf{b}_1 & \cdots & \mathbf{c}_R \otimes_{\mathrm{Kro}} \mathbf{b}_R \end{bmatrix} .$$

Finally, the Hadamard product of two matrices (of same dimensions) is the matrix given by the point-wise product of components, i.e, if A, B have matrix elements a_{ij} and b_{ij} then the Hadamard product $A \circ B$ has matrix elements $a_{ij}b_{ij}$.

Let $\mathbf{b}, \mathbf{d} \in \mathbb{R}^I$ and $\mathbf{c}, \mathbf{e} \in \mathbb{R}^J$ be column vectors. Let $B, D \in \mathbb{R}^{I \times R}$ and $C, E \in \mathbb{R}^{J \times R}$ be four matrices. Check the following identities used in class:

$$(\mathbf{c} \otimes_{\mathrm{Kro}} \mathbf{b})^T = \mathbf{c}^T \otimes_{\mathrm{Kro}} \mathbf{b}^T ;$$
$$(\mathbf{e} \otimes_{\mathrm{Kro}} \mathbf{d})^T(\mathbf{c} \otimes_{\mathrm{Kro}} \mathbf{b}) = (\mathbf{e}^T\mathbf{c})(\mathbf{d}^T\mathbf{b}) ;$$
$$(E \otimes_{\mathrm{Khr}} D)^T(C \otimes_{\mathrm{Khr}} B) = (E^TC) \circ (D^TB) .$$