Homework 8 - graded
CS-526 Learning Theory

Note: The tensor product is denoted by $\otimes$. In other words, for vectors $\underline{a}, \underline{b}, \underline{c}$ we have that $\underline{a} \otimes \underline{b}$ is the square array $a^\alpha b^\beta$ where the superscript denotes the components, and $\underline{a} \otimes \underline{b} \otimes c$ 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: Moments of Gaussian mixture model (GMM)

Consider the following mixture of Gaussians (we look at the special case where all the covariance matrices are isotropic, equal to $\sigma^2 I_{D\times D}$):

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

where $\underline{x}, \underline{a}_i \in \mathbb{R}^D$ are column vectors and the weights $w_i \in (0, 1]$ satisfy $\sum_{i=1}^{K} w_i = 1$.

1) For $j \in [D]$, $\underline{e}_j$ is the j^{th} canonical basis vector of $\mathbb{R}^D$. Prove the following identities for the mean vector, the second moment matrix and the third moment tensor:

$$\mathbb{E}[\underline{x}] = \sum_{i=1}^{K} w_i \, \underline{a}_i \quad ;$$

$$\mathbb{E}[\underline{x}\, \underline{x}^T] = \sigma^2 I_{D\times D} + \sum_{i=1}^{K} w_i \, \underline{a}_i \underline{a}_i^T \quad ;$$

$$\mathbb{E}[\underline{x} \otimes \underline{x} \otimes \underline{x}] = \sum_{i=1}^{K} w_i \, \underline{a}_i \otimes \underline{a}_i \otimes \underline{a}_i + \sigma^2 \sum_{j=1}^{D} \sum_{i=1}^{K} w_i \big(\underline{a}_i \otimes \underline{e}_j \otimes \underline{e}_j + \underline{e}_j \otimes \underline{e}_j \otimes \underline{a}_i + \underline{e}_j \otimes \underline{a}_i \otimes \underline{e}_j\big) \, .$$

2) Let R be a $K \times K$ orthogonal (rotation) matrix. Define the matrix $\widetilde{R}$ whose entries are $\widetilde{R}_{ij} = \frac{1}{\sqrt{w_i}} R_{ij} \sqrt{w_j}$, as well as the transformed vectors

$$\underline{a}_i' = \sum_{j=1}^{K} \widetilde{R}_{ij} \underline{a}_j \, .$$

Show that the mixture of Gaussians

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

has the same second moment matrix as the previous one.

Problem 2: Examples of tensors and their rank

We recall that the rank of a tensor is the minimum possible number of terms in a decomposition of a tensor as a sum of rank-one tensors. Let $e_0^T = (1,0)$ and $e_1^T = (0,1)$. Consider the following second-order tensors (also called mode-2 or 2-way tensors):

$$
\begin{aligned}
B &= e_0 \otimes e_0 + e_1 \otimes e_1 \\
P &= e_0 \otimes e_0 + e_1 \otimes e_1 + e_0 \otimes e_1 + e_1 \otimes e_0 \\
E &= e_0 \otimes e_0 + e_1 \otimes e_1 + e_0 \otimes e_1
\end{aligned}
$$

as well as the third-order tensors (mode-3 or 3-way):

$$
\begin{aligned}
G &= e_0 \otimes e_0 \otimes e_0 + e_1 \otimes e_1 \otimes e_1 \\
W &= e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0 \ .
\end{aligned}
$$

1) Draw the two and three-dimensional multiarrays for all these tensors.

2) Determine the rank of each tensor (and justify your answer).

3) Let $\epsilon > 0$ and

$$D_\epsilon = \frac{1}{\epsilon}(e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) - \frac{1}{\epsilon} e_0 \otimes e_0 \otimes e_0$$

Check that $\lim_{\epsilon \to 0} D_\epsilon = W$. In other words, the rank-3 tensor W can be obtained as a limit of a sum of two rank-one tensors: W is on the "boundary" of the space of rank-2 tensors.

Problem 3: Frobenius norm minimizations: matrix versus tensors.

The Frobenius norm $\| \cdot \|_F$ of a tensor is defined as the Euclidean norm of the multi-array:

$$\|T\|_F^2 = \sum_{\alpha,\beta,\gamma} |T^{\alpha\beta\gamma}|^2 \ .$$

We recall the following important theorem for matrices.

Theorem 1 (Eckart-Young-Mirsky theorem)**.** *Let $A \in \mathbb{C}^{M\times N}$ be a rank-R matrix whose singular value decomposition is given by $U\Sigma V^*$ where $U \in \mathbb{C}^{M\times M}$, $V \in \mathbb{C}^{N\times N}$ are both unitary matrices and $\Sigma \in \mathbb{R}^{M\times N}$ is a diagonal matrix with real nonnegative diagonal entries. Without loss of generality we assume that the singular values are arranged in decreasing order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{M,N\}}$ (where $\sigma_i = \Sigma_{ii}$). Then, the best rank-k ($k \leq R$) approximation of A is given by the truncated SVD $\hat{A} = U\widetilde{\Sigma}V^*$ with $\widetilde{\Sigma}$ the diagonal matrix whose diagonal entries are $\widetilde{\Sigma}_{ii} = \sigma_i$ if $1 \leq i \leq k$, $\widetilde{\Sigma}_{ii} = 0$ otherwise. More precisely:*

$$\|A - \hat{A}\|_F = \min_{S:\mathrm{rank}(S)\leq k} \|A - S\|_F \ .$$

1) Do you think the analogous problem for tensors is well posed? In other words, given a tensor T of order $p \geq 3$ and rank R, can we always find a order-p tensor $\widehat{T}$ whose rank is strictly smaller than R and that achieves the minimum of $\|T - S\|_F$ over all the order-p tensors S of rank $k < R$?

2) We wish to come back to the interesting phenomenon observed in question 4) of Problem 2. In this question we saw that an order-3 rank-3 tensor could be obtained as the limit of a sequence of rank-2 tensors. Use the Eckart-Young theorem to show that a rank $R+1$ matrix cannot be obtained as a limit of a sum of R rank-one matrices. Can we obtain a rank-$(R-1)$ matrix as the limit of a sequence of rank-R matrices? And what about tensors?

3) *Independent question on the Frobenius norm.* The multilinear transformation of a tensor is the new tensor $T(R_1, R_2, R_3)$ with components

$$T(R_1, R_2, R_3)^{\alpha\beta\gamma} = \sum_{\delta,\epsilon,\zeta} R_1^{\alpha\delta} R_2^{\beta\epsilon} R_3^{\gamma\zeta} T^{\delta\epsilon\zeta} \ .$$

Check that if R_1, R_2, R_3 are unitary matrices then the Frobenius norm is invariant, i.e., $\|T\|_F = \|T(R_1, R_2, R_3)\|_F$. You can limit your proof to real-valued tensors if you wish[1].

Problem 4: Kronecker and Khatri-Rao products

The *Kronecker product* $\otimes_{\mathrm{Kro}}$ of two vectors $\underline{a} \in \mathbb{R}^{I_1}$ and $\underline{b} \in \mathbb{R}^{I_2}$ is a vectorization of the tensor (or outer) product. This amounts to take the $I_1 \times I_2$ array $a^\alpha b^\beta = (\underline{a} \otimes \underline{b})^{\alpha\beta}$ and view it as a vector of size $I_1 I_2$. More precisely, we define the Kronecker product as the column vector:

$$\underline{a} \otimes_{\mathrm{Kro}} \underline{b} = \begin{bmatrix} a^1 \underline{b}^T & \cdots & a^{I_1} \underline{b}^T \end{bmatrix}^T \in \mathbb{R}^{I_1 I_2} \ .$$

Let $A = \begin{bmatrix} \underline{a}_1 & \cdots & \underline{a}_R \end{bmatrix}$ and $B = \begin{bmatrix} \underline{b}_1 & \cdots & \underline{b}_R \end{bmatrix}$ be matrices of dimensions $I_1 \times R$ and $I_2 \times R$. We define the *Khatri-Rao* product as the $I_1 I_2 \times R$ matrix

$$A \odot_{\mathrm{KhR}} B = \begin{bmatrix} \underline{a}_1 \otimes_{\mathrm{Kro}} \underline{b}_1 & \cdots & \underline{a}_R \otimes_{\mathrm{Kro}} \underline{b}_R \end{bmatrix} \ .$$

1) Assume that both A and B are full column rank. Prove that the Khatri-Rao product $A \odot_{\mathrm{KhR}} B$ is also full column rank.

Problem 5: Check useful identities

Besides the tensor product (ntroduced in class), and the Kronecker and Khatri-Rao products (introduced in the previous exercise), we also introduce the Hadamard product. The Hadamard product of two matrices (of same dimensions) is the matrix given by the pointwise 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 $a_{ij} b_{ij}$.

[1] In this case R_1, R_2, R_3 are orthogonal matrices.

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 that will be used in class:

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