Homework 7 (Graded): due Monday, May 13, 2024
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 $\underline{a} \otimes \underline{b} \otimes \underline{c}$ 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:

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

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

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

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:*

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

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?

Problem 4: Tensors

1) Consider the tensor $M = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

(a) Write down the matrix components of M.

(b) For the matrix M of part (a) exhibit an uncountable number of decompositions of the form $M = \vec{a} \otimes \vec{b} + \vec{c} \otimes \vec{d}$ using the rotation matrices

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \theta \in \mathbb{R}\,.$$

2) Consider the following tensor decomposition

$$T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Is this decomposition unique? Justify your answer. What is the rank of T?

3) Let $\vec{a}_1, \vec{a}_2 \in \mathbb{R}^2$ be linearly independent and $\vec{b}_1, \vec{b}_2 \in \mathbb{R}^2$ be linearly independent as well. We define $T = \vec{a}_1 \otimes \vec{b}_1 \otimes \vec{c} + \vec{a}_2 \otimes \vec{b}_2 \otimes \vec{c}$ where $\vec{c} \in \mathbb{R}^2$ is not the zero vector.

(a) Does Jennrich's theorem apply?

(b) Prove that the tensor rank of T is 2.