Exercise 8 (Graded homework 4): 20^{th} of May 2019
CS-526 Learning Theory

The deadline is Tuesday, May 28, 2019. Please hand in your homework during the lecture (May 27) or the exercise session (May 28). **No scan of handwritten homework is accepted**.

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 GMM

Consider the mixture of Gaussians model:

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

where $\underline{x}, \underline{a}_i \in \mathbb{R}^D$ are column vectors and the weights $w_i \in (0, 1]$ satisfy $\sum_{i=1}^K w_i = 1$. Here we look at the special case where all the covariance matrices are equal to $\sigma^2 I_{D\times D}$.

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 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. Give the matricizations of G and W along the three modes, i.e., give the matrices $G_{(n)}$ and $W_{(n)}$ for the three modes $n = 1, 2, 3$.[1]

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

3) Let O be an orthogonal 2×2 matrix. Show that $B = (Oe_0) \otimes (Oe_0) + (Oe_1) \otimes (Oe_1)$. Does Jennrich's theorem allow to conclude that a similar result is possible or is not possible for G? And what about W?

4) 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 ($k \leq R$)*

[1] See section 3.2 in Review of Rabanser, Shchur, Gunnemann on the course webpage.

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 $1 \le i \le k$, $\widetilde{\Sigma}_{ii} = 0$ otherwise. More precisely:

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

1) Do you think the analogous problem for tensors is well posed? In other words, given a tensor T of order $p \ge 3$ 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) Now we wish to come back to the interesting phenomenon observed in question 4) of Problem 2 where 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.* Recall that 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 rotation (orthogonal) 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.

Problem 4: Kronecker and Khatri-Rao products

The *Kronecker product* $\otimes_{Kro}$ 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 an $I_1 I_2$ vector. More precisely we define the Kronecker product as the $I_1 I_2$ column vector:

$$\underline{a} \otimes_{Kro} \underline{b} = (a^1 \underline{b}^T, \cdots, a^{I_1} \underline{b}^T)^T \ .$$

Let $A = [\underline{a}_1, \cdots, \underline{a}_R]$ and $B = [\underline{b}_1, \cdots, b_R]$ matrices of dimensions $I_1 \times R$ and $I_2 \times R$. We define the *Khatri-Rao* product as the new $I_1 I_2 \times R$ matrix

$$A \odot_{KhR} B = [\underline{a}_1 \otimes_{Kro} \underline{b}_1, \cdots, \underline{a}_R \otimes_{Kro} \underline{b}_R] \ .$$

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

2) Explain in detail in which step of Jennrich's algorithm[2] this fact is used.

[2] See section 4.1.1 in Review of Rabanser, Shchur, Gunnemann on the course webpage.

Problem 5: Jennrich's type algorithm for order 4 tensors

Consider an order four tensor

$$T = \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r \otimes \underline{d}_r$$

where $A = [\underline{a}_1, \cdots, \underline{a}_R] \in \mathbb{R}^{I_1 \times R}$, $B = [\underline{b}_1, \cdots, \underline{b}_R] \in \mathbb{R}^{I_2 \times R}$, $C = [\underline{c}_1, \cdots, \underline{c}_R] \in \mathbb{R}^{I_3 \times R}$ and $D = [\underline{d}_1, \cdots, \underline{d}_R] \in \mathbb{R}^{I_4 \times R}$ are full column rank.

1) Check that you can apply Jennrich's algorithm to a "flattened" version of T, namely the order three tensor

$$\widetilde{T} = \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes (\underline{c}_r \otimes_{Kro} \underline{d}_r) \ .$$

where Kro is the Kronecker product defined in the previous question.

2) Deduce that the rank R and A, B, C, D can be uniquely determined from the four-dimensional array of numbers $T^{\alpha\beta\gamma\delta}$ (up to permutations and scalings).

Problem 6: The Moore-Penrose pseudoinverse

Consider a $M \times N$ matrix $A \in \mathbb{C}^{N\times M}$. Its transpose and complex conjugate (also called Hermitian conjugate) is the $N \times M$ matrix $\bar{A}^T$ that we denote A^*. Let $A^\dagger \in \mathbb{C}^{N\times M}$ satisfy the following four conditions:

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A \ .$$

A theorem of Moore and Penrose states that such a matrix always exists and is unique. This matrix is called the Moore-Penrose pseudoinverse. Answer the following questions:

1) Let $\Sigma \in \mathbb{C}^{M\times N}$ be a diagonal matrix, that is, $\forall i \neq j : \Sigma_{ij} = 0$ (but you don't necessarily have $M = N$). Show that $\Sigma^\dagger$ is the $N \times M$ diagonal matrix with diagonal entries

$$\forall i \in \{1, 2, \ldots, \min\{M, N\}\} : (\Sigma^\dagger)_{ii} = \begin{cases} 1/\Sigma_{ii} & \text{if} \quad \Sigma_{ii} \neq 0 \ ; \\ 0 & \text{otherwise.} \end{cases}$$

2) Let $A = U\Sigma V^*$ be the singular value decomposition (SVD) of A, that is, both $U \in \mathbb{C}^{M\times M}$ and $V \in \mathbb{C}^{N\times N}$ are unitary matrices and $\Sigma \in \mathbb{R}^{M\times N}$ is a diagonal matrix with real nonnegative diagonal entries (the singular values). Give for $A^\dagger$ an expression that only involves U, V (or their inverse U^*, V^*) and $\Sigma^\dagger$.

3) Show that if A has full column rank then $A^\dagger = (A^*A)^{-1}A^*$ and $A^\dagger A = I_{N\times N}$.

4) Show that if A has full row rank then $A^\dagger = A^*(AA^*)^{-1}$ and $AA^\dagger = I_{M\times M}$.

5) Show that if A is a square matrix with full rank then $A^\dagger = A^{-1}$ is the usual inverse.

6) Let A have full column rank and B have full row rank. Check that $(AB)^\dagger = B^\dagger A^\dagger$.