Solutions to Homework 7
CS-526 Learning Theory

Problem 1: Moments of Gaussian mixture model (GMM)

1) For every $i \in [K]$, $\underline{d}_i$ is the i^{th} canonical basis vector of $\mathbb{R}^K$ and we define the latent random vector $\underline{h} \in \{\underline{d}_i : i \in [K]\}$ whose distribution is $\forall i \in [K] : \mathbb{P}(\underline{h} = \underline{d}_i) = w_i$. Finally, let $\underline{x} = \sum_{i=1}^K h_i \underline{a}_i + \underline{z}$ where $\underline{z} \sim \mathcal{N}(0, \sigma^2 I_{D\times D})$ is independent of $\underline{h}$. The random vector $\underline{x}$ has a probability density function $p(\cdot)$. We have:

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

$$\begin{aligned}\mathbb{E}[\underline{x}\underline{x}^T] &= \mathbb{E}[\underline{z}\underline{z}^T] + \sum_{i=1}^K \mathbb{E}[h_i]\underbrace{\mathbb{E}[\underline{z}]}_{=0}\underline{a}_i^T + \mathbb{E}[h_i]\underline{a}_i\mathbb{E}[\underline{z}]^T + \sum_{i,j=1}^K \underbrace{\mathbb{E}[h_i h_j]}_{=w_i\delta_{ij}}\underline{a}_i\underline{a}_j^T \\ &= \sigma^2 I_{D\times D} + \sum_{i=1}^K w_i\, \underline{a}_i\underline{a}_i^T \ .\end{aligned}$$

Finally, to compute the third moment tensor, note that $\mathbb{E}[\underline{z}\otimes\underline{z}\otimes\underline{z}] = 0$ and that for every $(i,j) \in [K]^2$: $\mathbb{E}[\underline{a}_i \otimes \underline{a}_j \otimes \underline{z}] = \mathbb{E}[\underline{a}_i \otimes \underline{z} \otimes \underline{a}_j] = \mathbb{E}[\underline{z} \otimes \underline{a}_i \otimes \underline{a}_j] = 0$. Hence:

$$\begin{aligned}\mathbb{E}[\underline{x}\otimes\underline{x}\otimes\underline{x}] = &\sum_{i,j,k=1}^K \underbrace{\mathbb{E}[h_i h_j h_k]}_{=w_i\delta_{ij}\delta_{ik}}\underline{a}_i\otimes\underline{a}_j\otimes\underline{a}_k \\ &+ \sum_{i=1}^K \mathbb{E}[h_i]\mathbb{E}[\underline{a}_i\otimes\underline{z}\otimes\underline{z}] + \mathbb{E}[h_i]\mathbb{E}[\underline{z}\otimes\underline{a}_i\otimes\underline{z}] + \mathbb{E}[h_i]\mathbb{E}[\underline{z}\otimes\underline{z}\otimes\underline{a}_i] \\ = &\sum_{i=1}^K w_i\, \underline{a}_i\otimes\underline{a}_i\otimes\underline{a}_i + \sigma^2\sum_{j=1}^D\sum_{i=1}^K w_i(\underline{a}_i\otimes\underline{e}_j\otimes\underline{e}_j + \underline{e}_j\otimes\underline{e}_j\otimes\underline{a}_i + \underline{e}_j\otimes\underline{a}_i\otimes\underline{e}_j)\,.\end{aligned}$$

2) Let $A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_K] \in \mathbb{R}^{D\times K}$ and $A' = [\underline{a}'_1, \underline{a}'_2, \ldots, \underline{a}'_K] \in \mathbb{R}^{D\times K}$. By definition, $\widetilde{R} = \Sigma^{-1}R\Sigma$ where Σ is the diagonal matrix such that $\Sigma_{ii} = \sqrt{w_i}$ and $A' = A\widetilde{R}^T$. We can directly apply the formula of question 1) to compute the second moment matrix of the new mixture of Gaussians:

$$\begin{aligned}\mathbb{E}[\underline{x}\underline{x}^T] &= \sigma^2 I_{D\times D} + A'\Sigma^2 A'^T = \sigma^2 I_{D\times D} + A\widetilde{R}^T\Sigma^2\widetilde{R}A^T \\ &= \sigma^2 I_{D\times D} + A\Sigma R^T R\Sigma A^T = \sigma^2 I_{D\times D} + A\Sigma^2 A^T\ .\end{aligned}$$

Problem 2: Examples of tensors and their rank

1) The matrices corresponding to B, P, E are:

$$B = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\ ;\ P = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\ ;\ E = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$$

The frontal slices of G and W are:

$$G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} , \; G_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \; ; \; W_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} , \; W_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} .$$

2) B and E are clearly rank-2 matrices, while $P = (e_0 + e_1) \otimes (e_0 + e_1)$ is a rank-1 matrix.
By its definition, G is at most rank 2. Assume it is rank 1: $G = a \otimes b \otimes c$ with $a, b, c \in \mathbb{R}^2$. We have $a_1 b_1 c_1 = G_{111} = 1$ and $a_2 b_1 c_1 = G_{211} = 0$ so we must have $a_2 = 0$. Besides, $a_2 b_2 c_2 = G_{222} = 1$ and $a_1 b_2 c_2 = G_{122} = 0$ so $a_1 = 0$. Hence $a^T = (0, 0)$ and G is the all-zero tensor. This is a contradiction and we conclude that G is rank 2.
By its definition, W is at most rank 3. To prove the rank cannot be smaller than 3, we will proceed by contradiction:

- Assume W is rank 1: $W = a \otimes b \otimes c$ with $a, b, c \in \mathbb{R}^2$. We have $a_1 b_1 c_1 = W_{111} = 0$ and $a_2 b_1 c_1 = W_{211} = 1$ so $a_1 = 0$. Besides, $a_1 b_1 c_2 = W_{112} = 1$ and $a_2 b_1 c_2 = W_{212} = 0$ so $a_2 = 0$. Then $a = (0, 0)^T$ and W is the all-zero tensor, which is a contradiction.

- Assume W is rank 2: $W = a \otimes b \otimes c + d \otimes e \otimes f$. We claim that *$a$ and d must be linearly independent*. Indeed, suppose they are parallel and take a vector x perpendicular to both a and d. Then
$$W(x, I, I) = (x^T a) b \otimes c + (x^T d) e \otimes f = 0$$
but also
$$W(x, I, I) = (x^T e_0) e_0 \otimes e_1 + (x^T e_0) e_1 \otimes e_0 + (x^T e_1) e_0 \otimes e_0 = \begin{bmatrix} x^T e_1 & x^T e_0 \\ x^T e_0 & 0 \end{bmatrix}$$
which cannot be zero since x cannot be perpendicular to both e_0 and e_1. Now, we take x perpendicular to d. We have
$$W(x, I, I) = (x^T a) b \otimes c$$
which is rank one. Therefore, we must have $x^T e_0 = 0$ which implies that x is parallel to e_1 and thus <u>d parallel to e_0</u>. Now, if we take x perpendicular to a, the matrix
$$W(x, I, I) = (x^T d) e \otimes f$$
is rank one and, once again, we must have $x^T e_0 = 0$, which implies x parallel to e_1 and thus <u>a parallel to e_0</u>. Hence, we have shown that a and d are linearly independent but also that both are parallel to e_0. This is a contradiction.

3) We expand the tensor products in the definition of D_ϵ:

$$\begin{aligned} D_\epsilon &= \frac{1}{\epsilon} \Big[(e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) - e_0 \otimes e_0 \otimes e_0 \Big] \\ &= \frac{1}{\epsilon} \Big[e_0 \otimes e_0 \otimes e_0 + \epsilon\, e_0 \otimes e_0 \otimes e_1 + \epsilon\, e_0 \otimes e_1 \otimes e_0 + \epsilon\, e_1 \otimes e_0 \otimes e_0 \\ &\qquad + \epsilon^2\, e_1 \otimes e_1 \otimes e_0 + \epsilon^2\, e_1 \otimes e_0 \otimes e_1 + \epsilon^2\, e_0 \otimes e_1 \otimes e_1 + \epsilon^3\, e_1 \otimes e_1 \otimes e_1 - e_0 \otimes e_0 \otimes e_0 \Big] \\ &= e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0 \\ &\qquad + \epsilon (e_1 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_1) + \epsilon^2\, e_1 \otimes e_1 \otimes e_1 \\ &= W + \epsilon (e_1 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_1) + \epsilon^2\, e_1 \otimes e_1 \otimes e_1 \; . \end{aligned}$$

Hence $\lim_{\epsilon \to 0} D_\epsilon = W$.

Problem 3: Frobenius norm minimizations: matrix versus tensors.

1) There cannot be an analogous general result for tensors. Indeed, the order-3 tensor W of Problem 2 is rank 3 and we showed in 3) that $\lim_{\epsilon \to 0} \|W - D_\epsilon\|_F = 0$. So there is no minimum attained in the space of rank 2 tensors. In this sense, there is simply no *best* rank-two approximation of W.

2) Let M a matrix of rank $R+1$ with singular values $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_R \geq \sigma_{R+1} > 0$. By the Eckart-Young-Mirsky theorem, the minimum of $\|M - \widehat{M}\|_F$ over all the matrices $\widehat{M}$ of rank less than, or equal to, R is $\sigma_{R+1} > 0$. Therefore, there cannot be a sequence of matrices M_n given by a sum of R rank-one matrices such that $\lim_{n\to+\infty} \|M - M_n\|_F = 0$.

Now let $M \in \mathbb{C}^{M\times N}$ be a matrix of rank $R-1$ with $R \leq \min\{M, N\}$. Let $M = U\Sigma V^*$ be the SVD of M where $\sigma_1 \geq \cdots \geq \sigma_{R-1} > 0$ are its singular values. For all positive integer n, we define $\sigma_R^{(n)} := \sigma_{R-1}/n$ as well as the rank-R matrix $M_n = U\Sigma_n V^*$ where Σ_n is a $M \times N$ diagonal matrix whose nonzero diagonal entries are $\sigma_1 \geq \cdots \geq \sigma_{R-1} \geq \sigma_R^{(n)}$. Clearly $\lim_{n\to+\infty} \|M - M_n\|_F = \lim_{n\to+\infty} \frac{\sigma_{R-1}}{n} = 0$. A similar procedure can be applied if M is a tensor.

Problem 4: Tensors

1. (a) $M = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$.

   (b) Using that $RR^T = I_2$:

$$M = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} R \cdot R^T \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \vec{a} \otimes \vec{b} + \vec{c} \otimes \vec{d}$$

   where

$$\begin{bmatrix} \vec{a} & \vec{c} \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} R = \begin{pmatrix} \cos\theta + \sin\theta & \cos\theta - \sin\theta \\ 2\cos\theta & -2\sin\theta \end{pmatrix} ;$$
$$\begin{bmatrix} \vec{b} & \vec{d} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \cos\theta + \sin\theta & \cos\theta - \sin\theta \end{pmatrix} .$$

2. We have $T = \vec{a}_1 \otimes \vec{b}_1 \otimes \vec{c}_1 + \vec{a}_2 \otimes \vec{b}_2 \otimes \vec{c}_2 + \vec{a}_3 \otimes \vec{b}_3 \otimes \vec{c}_3$ where

$$\begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 1 & 2 & 5 \end{bmatrix} \text{ has pairwise independent columns;}$$
$$\begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ has linearly independent columns;}$$
$$\text{and } \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \vec{c}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ has linearly independent columns.}$$

   By Jennrich's theorem the decomposition is therefore unique and the rank of T is 3.

3. (a) We have $T = \vec{a}_1 \otimes \vec{b}_1 \otimes \vec{c}_1 + \vec{a}_2 \otimes \vec{b}_2 \otimes \vec{c}_2$ where $\vec{c}_1 = \vec{c}_2 = \vec{c}$. We cannot invoke Jennrich's theorem because the vectors $\vec{c}_1, \vec{c}_2$ are not pairwise independent.

(b) The tensor rank is obviously less than or equal to 2. We will prove by contradiction that it cannot be equal to 1.

Assume the rank is one. Then there exist vectors $\vec{e}, \vec{f}, \vec{g}$ such that $T = \vec{e} \otimes \vec{f} \otimes \vec{g}$. Pick any vector $\vec{x}$ that is not orthogonal to $\vec{c}$. We have:

$$(\vec{e} \otimes \vec{f})(\vec{g}^T \vec{x}) = (\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2)(\vec{c}^T \vec{x})$$

The matrix $(\vec{e} \otimes \vec{f})(\vec{g}^T \vec{x})$ has rank 0 or 1 while the matrix $(\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2)(\vec{c}^T \vec{x})$ has rank 2 because $\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2$ has rank 2 and $\vec{c}^T \vec{x} \neq 0$. This is a contradiction.