Solutions to Homework 10
CS-526 Learning Theory

Problem 1: Multilinear Rank, Tensor Rank

Recall the formulas for the matricizatons:
$T_{(1)} = A(C \otimes_{KhR} B)^T$, where A and $C \otimes_{KhR} B$ are of dimensions $I_1 \times R$ and $I_2 I_3 \times R$ respectively. Moreover for any matrices X, Y we have that:

$$\text{rank}(XY) \leq \min\{\text{rank}(X), \text{rank}(Y)\}$$

Thus:

$$R_1 = \text{rank}(T_{(1)}) \leq \text{rank}(A) \leq \min\{I_1, R\} \leq R$$

By repeating the same argument for matrizications $T_{(2)}$, $T_{(3)}$ we conclude the proof.

Problem 2: Non-unicity of Tucker decomposition

Let $X = [\vec{x_1}, \dots, \vec{x_{R_1}}]$, $Y = [\vec{y_1}, \dots, \vec{y_{R_2}}]$, and $Z = [\vec{z_1}, \dots, \vec{z_{R_3}}]$. Then, from the definitions of vectors $\vec{x_{p'}}$, $\vec{y_{q'}}$, $\vec{z_{r'}}$, and from the orthogonality of the matrices $M^{(u)}, M^{(v)}, M^{(w)}$ it is easy to see that:

1. $U \cdot (M^{(u)})^T = X \Rightarrow U = X \cdot M^{(u)} \Rightarrow \vec{u}_p = X \cdot M^{(u)}_{:p} = \sum_{p'} M^{(u)}_{p'p} \vec{x}_{p'}$
2. $V \cdot (M^{(v)})^T = Y \Rightarrow V = Y \cdot M^{(v)} \Rightarrow \vec{v}_q = Y \cdot M^{(v)}_{:q} = \sum_{q'} M^{(v)}_{q'q} \vec{y}_{q'}$
3. $W \cdot (M^{(w)})^T = Z \Rightarrow W = Z \cdot M^{(w)} \Rightarrow \vec{w}_r = Z \cdot M^{(w)}_{:r} = \sum_{r'} M^{(w)}_{r'r} \vec{z}_{r'}$

Substituting $\vec{u}_p, \vec{v}_q$ and $\vec{w}_r$ in the Tucker decomposition expression we get:

$$T = \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr}\, \vec{u}_p \otimes \vec{v}_q \otimes \vec{w}_r = \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr}\, (\sum_{p'} M^{(u)}_{p'p} \vec{x}_{p'}) \otimes (\sum_{q'} M^{(v)}_{q'q} \vec{y}_{q'}) \otimes (\sum_{r'} M^{(w)}_{r'r} \vec{z}_{r'}) =$$
$$\sum_{p',q',r'=1}^{R_1,R_2,R_3} \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr} M^{(u)}_{p'p} M^{(v)}_{q'q} M^{(w)}_{r'r}\, \vec{x}_{p'} \otimes \vec{y}_{q'} \otimes \vec{z}_{r'} = \sum_{p',q',r'=1}^{R_1,R_2,R_3} H^{p'q'r'}\, \vec{x}_{p'} \otimes \vec{y}_{q'} \otimes \vec{z}_{r'}$$

where $H^{p'q'r'} = \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr} M^{(u)}_{p'p} M^{(v)}_{q'q} M^{(w)}_{r'r}$, which concludes the proof.

Problem 3: Whitening of a tensor

1. We have $M = U\text{Diag}(d_1, \dots, d_K)U^T$ and, by definition, $W := U\text{Diag}(d_1^{-1/2}, \dots, d_K^{-1/2})$. A direct computation gives:

$$\begin{aligned} W^T M W &= \text{Diag}(d_1^{-1/2}, \dots, d_K^{-1/2})(U^T U)\text{Diag}(d_1, \dots, d_K)(U^T U)\text{Diag}(d_1^{-1/2}, \dots, d_K^{-1/2}) \\ &= \text{Diag}(d_1^{-1/2}, \dots, d_K^{-1/2})\text{Diag}(d_1, \dots, d_K)\text{Diag}(d_1^{-1/2}, \dots, d_K^{-1/2}) \\ &= I\ . \end{aligned}$$

We used that the columns of U are orthogonal unit vectors: $U^TU = I$. By definition of $\vec{v}_i$, we have $V := \begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_K\end{bmatrix} = W^T \mu \,\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})$ where $\mu := \begin{bmatrix}\vec{\mu}_1 & \cdots & \vec{\mu}_K\end{bmatrix}$. It also follows from the definition of M that $M = \mu \,\mathrm{Diag}(\lambda_1, \ldots, \lambda_K)\mu^T$. Hence:

$$\begin{aligned} VV^T &= W^T \mu \,\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})\mu^T W \\ &= W^T M W \\ &= I\ . \end{aligned}$$

The matrix V is square and satisfies $VV^T = I$, thefore $V^TV = I$ meaning that the vector $\vec{v}_i$ are orthonormal.

2. Because M is known we can compute the matrix W and use it to obtained the whitened tensor $T(W, W, W) = \sum_{i=1}^K \nu_i \,\vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$ where $\nu_i = \lambda_i^{-1/2}$ and $\vec{v}_i = \sqrt{\lambda_i}\, W^T \vec{\mu}_i$. We have shown in the previous question that $\vec{v}_1, \ldots, \vec{v}_K$ are orthogonal unit vectors. Thus, we can use the tensor power method to recover each of the pair $\pm(\nu_i, \vec{v}_i)$ for $i \in [K]$. Because $\nu_i > 0$ we can disambiguate the sign and determine $(\nu_i, \vec{v}_i)$ from $\pm(\nu_i, \vec{v}_i)$.
Now that all the $(\nu_i, \vec{v}_i)$ are known, we need to show that the whitening transformation can be inverted to give back $(\lambda_i, \vec{\mu}_i)$. The relation between λ_i and ν_i is easy to invert: $\lambda_i = 1/\nu_i^2$. To recover $\mu = \begin{bmatrix}\vec{\mu}_1 & \cdots & \vec{\mu}_K\end{bmatrix}$, we need to invert the system of equations

$$V = W^T \mu \,\mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K}) \Leftrightarrow V\mathrm{Diag}(\nu_1, \ldots, \nu_K) = W^T \mu\ . \tag{1}$$

The matrix $W^T = \mathrm{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2})U^T$ has full row rank and its Moore-Penrose pseudo-inverse reads $(W^T)^\dagger = U\mathrm{Diag}(\sqrt{d_1}, \ldots, \sqrt{d_K})$. Multiplying both sides of (1) by $(W^T)^\dagger$ yields:

$$(W^T)^\dagger V\mathrm{Diag}(\nu_1, \ldots, \nu_K) = UU^T \mu\ . \tag{2}$$

At this point we might be tempted to say that $UU^T = I$, yielding $\mu = (W^T)^\dagger V\mathrm{Diag}(\nu_1, \ldots, \nu_K)$. However, U is in general not a square matrix and we cannot conclude $UU^T = I$ from $U^TU = I$. This is only a minor setback. Note that (the left-hand side is the definition of M, the right-hand side is its diagonalization):

$$\mu\mathrm{Diag}(\lambda_1, \ldots, \lambda_K)\mu^T = U\mathrm{Diag}(d_1, \ldots, d_K)U^T\ ,$$

where μ, U are $D \times K$ full column rank matrices. It follows that $\mathrm{span}(\mu) = \mathrm{span}(U)$ and there exists a $K \times K$ matrix P such that $\mu = UP$. Hence, $UU^T\mu = U(U^TU)P = UP = \mu$ and (2) reads:

$$\mu = (W^T)^\dagger V\mathrm{Diag}(\nu_1, \ldots, \nu_K) = U\mathrm{Diag}(\sqrt{d_1}, \ldots, \sqrt{d_K})V\mathrm{Diag}(\nu_1, \ldots, \nu_K)\ .$$

We are thus able to recover μ from the knowledge of W, V and $\mathrm{Diag}(\nu_1, \ldots, \nu_K)$.