ISYE 8803 HDDA

Homework 3
June 21s, 2021

Problem 2. CP and Tucker decomposition

a. To minimize the least square problem

$$\{C, \boldsymbol{V}_1, \dots, \boldsymbol{V}_d\} = \text{argmin}\left\|Y - \sum_{i=1}^{p} X_i B\right\|$$

$$B_j = C_j \times_1 U_{j1} \times_2 U_{j2} \times_3 \dots \times_{lj} U_{jl_j} \times_{l_j+1} V_1 \times_{l_j+2} \dots \times_{l_j+d} V_d$$

$$V_i^T V_i = I_{\tilde{Q}_i}$$

Where $Y \in R^{Q_1 \times \dots \times Q_d}$ and $X_i \in R^{P_{i1} \times \dots \times P_{il_i}}$ $C_j \in R^{\tilde{P}_1 \times \dots \times \tilde{P}_{lj} \times_{l_j+1} \tilde{Q}_1 \times_{l_j+2} \dots \times_{l_j+d} \tilde{Q}_d}$ is a core tensor with $\tilde{P}_{ji} \ll P_{ji}$ and $\tilde{Q}_i \ll Q_i$; $\{U_{ji}: j = 1, \dots, p; i = 1, \dots, l_j\}$ is a set of bases that spans the j^{th} input space; and $\{V_i: i = 1, \dots, d\}$ is a set of bases that spans the output space.

Prove the following theorem:

When U_{ji}, V_i and R_j are known, a reshape form of the core tensor C_j can be estimated as

$$\tilde{C}_j = R_j \times_1 \left(\boldsymbol{Z}_j^T \boldsymbol{Z}_j\right)^{-1} \boldsymbol{Z}_j^T \times_2 \boldsymbol{V}_1^T \times_3 \boldsymbol{V}_2^T \dots \times_{d+1} \boldsymbol{V}_d^T$$

Where $Z_j = X_{j(1)}(\boldsymbol{U}_{jl} \otimes \boldsymbol{U}_{jl-1} \otimes \dots \otimes \boldsymbol{U}_{j1})$ and $R_j = Y - \sum_{i \neq j}^{p} B_j * X_j$. Note that $\tilde{C}_j$ has fewer modes $(d+1)$ than the original core tensor in (4), but it can be transformed into C by a simple reshape operation.

This can be done by the following steps:

1) Prove $\text{argmin}_C \left\|R_{j(1)} - X_{j(1)} B_j\right\|_F^2 = \text{argmin}_C \left\|vec(R_{j(1)}) - (V_d \otimes V_{d-1} \dots \otimes V_1 \otimes Z_j) vec(C_j)\right\|_F^2$ where $vec(X)$ stacks the columns of matrix X on top of each other. $Hint: \left(vec(ABC^T) = (C \otimes A) vec(B)\right)$

a)

part 1)

We know from the above definitions that when we substiute B with it's expanded form we get the following:

$$argmin_c ||R_{j(1)} - X_{j(1)}(U_{jl} \bigotimes U_{jl-1} \dots \bigotimes U_{j1}) C_j (V_d \bigotimes V_{d-1} \dots \bigotimes V_1)||_F^2$$

We can see that $X_{j(1)}(U_{jl} \bigotimes U_{jl-1} \dots \bigotimes U_{j1})$ *can be substiuted for* Z_j *to get the following:*

We can apply the vectorization to R and C since it just changed the positioning of the *elements which yields the following* :

$argmin_c ||vec(R_{j(1)}) - (V_d \bigotimes V_{d-1} \ldots \bigotimes V_1 \bigotimes Z_j\ )vec(C_j)||_F^2$

part 2)

When we take the derivative of the previous result we get the following :

$$-2(V_d \bigotimes V_{d-1} \ldots \bigotimes V_1 \bigotimes Z_j\ )^T((vec(R_{j(1)}) - (V_d \bigotimes V_{d-1} \ldots \bigotimes V_1 \bigotimes Z_j\ )\ vec(C_j)) = 0$$

$$\left(V_d{}^T \bigotimes V_{d-1}{}^T \ldots \bigotimes V_1{}^T\right) \bigotimes Z_j))\ vec(R_{j(1)})) \ =$$
$$\left(V_d{}^T \bigotimes V_{d-1}{}^T \ldots \bigotimes V_1{}^T\right) \bigotimes Z_j{}^T))\ (V_d \bigotimes V_{d-1} \ldots \bigotimes V_1 \bigotimes Z_j\ )\ vec(C_j)$$

Then when using properties $(A \bigotimes B)^T = \ A^T \bigotimes B^T$, $(A \bigotimes B)^{-1} = \left(A^{-1} \bigotimes B^{-1}\right)$, *we can then yield* :

$$\left(V_d{}^T \bigotimes V_{d-1}{}^T \ldots \bigotimes V_1{}^T\right) \bigotimes \ \left(Z_j{}^T Z_j\right)^{-1} Z_j^T))\ vec(R_{j(1)})) \ = \ \ vec(C_j)$$

We can then prove that :

$$\tilde{C}_j = R_j \times_1 \left(\mathbf{Z}_j^T \mathbf{Z}_j\right)^{-1} \mathbf{Z}_j^T \times_2 \mathbf{V}_1^T \times_3 \mathbf{V}_2^T \ldots \times_{d+1} \mathbf{V}_d^T$$

b)

From the lectures we know that CP decomposition is a special case of Tucker decomposition because the G matrix yields a super diagonal where all of the column vectors are orthonormal and P=Q=R.

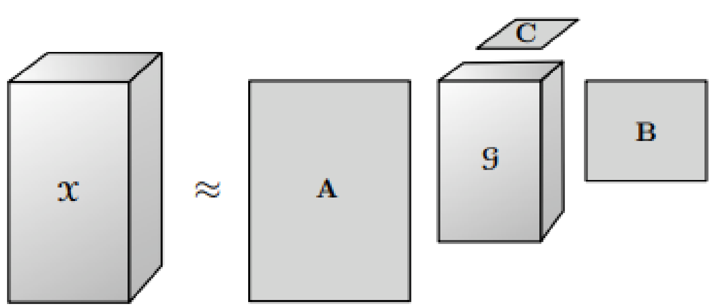

Figure 1: Tucker decomp