HW3 : DeepLearning , CSCE - 636

SIRI PRANITHA NAMBURI
331007601

① Given, the eigen value decomposition of A

$A = Q \Lambda Q^T$ where $Q = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix}$

$\Lambda = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$

SVD of A would be in form $A = U \tilde{\Sigma} V^T$;
U, V are orthogonal matrices, $\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$ $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$

Given A is a symmetric matrix, Q is orthogonal ~~and has~~

$$A = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

transforming the negative sign, by changing sign of relevant column in U.

$$A = \begin{bmatrix} u_{11} & u_{12} & -u_{13} \\ u_{21} & u_{22} & -u_{23} \\ u_{31} & u_{32} & -u_{33} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{12} & u_{22} & u_{32} \\ u_{13} & u_{23} & u_{33} \end{bmatrix}$$

transforming order of values to make sure diagonal elements of the middle matrix are in descending order.

$$A = \begin{bmatrix} u_{11} & -u_{13} & u_{12} \\ u_{21} & -u_{23} & u_{22} \\ u_{31} & -u_{33} & u_{32} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{32} \end{bmatrix} \quad — (1)$$

Hence equation (1) is the SVD of A with.

$$U = \begin{bmatrix} u_{11} & -u_{13} & u_{12} \\ u_{21} & -u_{23} & u_{22} \\ u_{31} & -u_{33} & u_{32} \end{bmatrix} \quad \tilde{\Sigma} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad V = \begin{bmatrix} u_{11} & u_{21} & u_{31} \\ u_{13} & u_{23} & u_{33} \\ u_{12} & u_{22} & u_{32} \end{bmatrix}$$

(2) PROOF OF KYFAN THEROM

$H \in \mathbb{R}^{n \times n}$; H is symmetric matrix with eigen values $\lambda_1 \geq \dots \geq \lambda_n$; corresponding eigen vectors $U = [u_1, \dots u_n]$

we have to prove that $\lambda_1 + \dots + \lambda_k = \max_{A \in \mathbb{R}^{n \times k};\, A^T A = I_k} \operatorname{trace}(A^T H A)$

and the optimal $A^* = [u_1, \dots, u_k] Q$; Q = arbitary ortrogonal matrix

proof: ~~Here~~ let eigen decomposition of H be

$H = U \Lambda U^T$; Λ = diagnol matrix with eigen values.

$$A^T H A = A^T (U \Lambda U^T) A = A^T U \Lambda U^T A.$$

assume $A^T U = B^T$. Then $\boxed{A^T H A = B^T \Lambda B}$

$$\operatorname{trace}(A^T H A) = \operatorname{trace}(B^T \Lambda B)$$

$$= \sum_{n,k,j} t_{nk} \Lambda_{kj} t_{jn}$$

$$= \sum_k \left(\Lambda_{kk} \sum_n t^T_{kn} t_{nk} \right)$$

Here $t^T_{kn} t_{nk} \leq 1$ as A is semiortrogonal

and U is ortrogonal

Here A is semiorthogonal as $A^TA = I_k$; i.e;
$A^T = A^{-1}$ for $k \times k$ rows and columns, with rest being 0.

$t_{kn}^T t_{nk} = 1$, is when $v_i a_i$ is not orthogonal for
$i = 0, 1, \ldots, k$ $\quad v_i^T a_i = 1$

and when $v_j a_j$ is orthogonal for $j = k+1, \ldots n$
$v_i^T a_i = 0$.

$$\max \operatorname{trace}(B^T \Lambda B) = \max \Big(\sum_{i=1}^{k} \lambda_i \sum_{i=1}^{k} b_i^T b_i + \sum_{j=k+1}^{n} \lambda_j \sum_{j=k+1}^{n} b_j^T b_j \Big)$$

$$= \sum_{i=1}^{k} \lambda_i = \lambda_1 + \cdots \lambda_k.$$

Hence $\lambda_1 + \cdots + \lambda_k = \max\limits_{A \in \mathbb{R}^{n \times k}, A^TA = I_k} \operatorname{trace}(A^T H A)$

$\forall i \in [1, k]$, $\|A^T u_i\|_2 = 1 \Rightarrow u_i = A q_i$ where
$q_1, \ldots q_k$ are orthonormal

For $k+1 \leq i \leq n \Rightarrow \|A u_i\| = 0$
Hence $A^* = [u_1, \ldots u_k] Q$ where
$Q = [q_1, \ldots q_k] \in \mathbb{R}^{k \times k}$

Scanned with CamScanner

5) Bonus Question.

Let us take the training and testing kernels as K_1 and K_2 $K_1 \in \mathbb{R}^{n \times n}$, $K_2 \in \mathbb{R}^{n \times m}$.

training data: $A \in \mathbb{R}^{n \times k}$; testing data: $B \in \mathbb{R}^{m \times k}$ for some number of features k.

$$K_1 = AA^T \text{ ; let compute SVD of } K_1 = U \Sigma U^T$$

$$K_1 = U \Sigma^{1/2} (U \Sigma^{1/2})^T$$ ($\because \Sigma$ has positive values on the diagonal)

$$\therefore \boxed{A = U \Sigma^{1/2}}$$

Hence A, aka, the training data can be computed using U and Σ and it can be used for training primal solvers.

For testing, $K_2 = AB^T \Rightarrow K_2 = U \Sigma^{1/2} B^T$

$$U^T K_2 = U^T U \Sigma^{1/2} B^T$$ ($U^T U = I$ as U is orthogonal)

$$U^T K_2 = \Sigma^{1/2} B^T$$

$$B^T = (\Sigma^{1/2})^{-1} U^T K_2$$

$$\boxed{B = (\Sigma^{-1/2} U^T K_2)^T}$$

In this way, the testing data B can also be computed. We use this to get predictions.