Mohammad Alshurbaji

Adv. ML

Assignment 2.

Problem 1: Section 1:

Show that the first column of U.

$svd(X^T X) = U S$

where U is matrix of orthonormal eigenvectors of $X^T X$.

if $\|X\Phi\|_2^2$ will be max. from first column of U.

Note that: $\|\Phi\|_2^2 = \text{Trace}(\Phi^T \Phi)$ ..①

so, max. $\|X\Phi\|_2^2$, s.t. $\|\Phi\|_2 = 1$ is:

$$\max.\ \text{Trace}(X^T \Phi^T \Phi X) \quad \text{s.t. } \|\Phi_n\| = 1 \quad ..②$$

Note that: $\Phi^T \Phi = I$ and $X^T X$ is the covariance matrix.

Now: let's take eigen value:

$$X^T X \Phi - \lambda \Phi$$

where λ is the eigenvalue of $X^T X$:

$$L(\theta, \lambda) = \theta^T X^T X \theta - \lambda(\theta^T \theta - 1)$$

$$= \theta^T X^T X \theta - \lambda \theta^T \theta + \lambda$$

max: $$\frac{\partial L}{\partial \theta} = 2 X^T X \theta - 2\lambda\theta = 0$$

$$\therefore \theta^T X^T X \theta - \lambda\theta - \theta^T \lambda \theta = \lambda \|\theta\|_2 = \lambda$$

which shows that the value of λ will be the largest (first column of U). #

Finally:

$$\hat{\Phi} = 0.8 * u_1 + 0.6 u_2 \quad , \quad \|\Phi\|_2^2 = 1$$

$$\|0.8 u_1 + 0.6 u_2\|^2 = \|0.8 u_1\|^2 + \|0.8 u_2\|^2 + 2 \langle 0.8 u_1 + 0.8 u_2 \rangle$$

$$= 0.64 \|u_1\|^2 + 0.36 \|u_1\|^2 + 0$$

problem $\cancel{2}$ 1: section 2

note that:

$$\|X\phi\|_2^2 = (X\phi)^T(X\phi) = \phi^T(X^TX)\phi$$

what if: $\|X(-\phi)\|_2^2 = \cancel{\phi}\,(-\phi)^T(X^TX)(-\phi)$

$$\equiv \phi^T(X^TX)\phi$$

$$\equiv \|X\phi\|_2^2$$

By this we could say no effect for the Negativity and the sol. is not Unique.

section 3

Problem $\cancel{2}$ 1: Show that first r columns of U max. $\|XW\|_F^2$ s.t. $W^TW = I_r$

$$\|XW\|_F^2 = Tr(X^TW^TXW) = \sum_{i=1}^{r} \sigma_i^2$$

Then: $\max\limits_{W^TW=I_r} tr(W^TXX^TW)$

So, By choosing W as the first r columns of U, we ensure that W spans the principal components associated with the largest r singular values of X. and this max. the sum $\sum_{i=1}^{r}\sigma_i^2$.

problem $\cancel{4}$ 1: section 4.

The possible different W's will max. the obj. is : $\boxed{2^r}$ (2 because of $\phi, -\phi$).

Problem 2:

Given $X \in R^{n \times p}$, already centered

$n \rightarrow$ Sample Size

$p \rightarrow$ Feature Size

Normal Method: finding eigenvectors to the largest eigenvalues of $X^T X$. comp. $O(P^3)$

Fast approach to reduce complexity:

PCA based on $X X^T$ and obtain eigenvectors for $X^T X$ accordingly.

Note: eigenvalues of $X^T X$ and $X X^T$ are the same.

Note: eigvec. of $X^T X$ is:

$\boxed{X^T * \text{eig vect. of } X X^T}$

After conducting an experiment using Matlab:

[100,1000] For $n \ll P \rightarrow$ The Fast approach is much faster than the Normal approach. While, for small matrices, there is no difference.

problem 3:

$\theta^* \in R^d$ the ground Truth Lin. Mod. Param.

$X \in R^{N \times d}$ The observing matrix, each column is Indepentent.

Linear Model: $y = X\theta^* + \epsilon$; ϵ follows Gaussian $(0, \sigma^2 I)$

$\hat{\theta} = \arg\min_{\theta} \|X\theta - y\|^2$

Section 1: show that $X^T X$ is Invertible.

$X \in R^{N \times d} \rightarrow X^T X \in R^{d \times d}$ ...①

As It's Linearly Independent since each col. of X is Indepentant, means the rank of X is d.

And as $X^T X \in R^{d \times d}$ → means It's Full Rank

$\underbrace{d}_{\text{Rows}} \times \underbrace{d}_{\text{columns}}$.

Finally, In order for a Matrix to be Invert., It should be Full Rank.

Here it is, It's Invertible. #

Section 2: Show that $MSE(\theta^*, \hat{\theta}) :=$

$$E_\epsilon \{ \| \theta^* - \hat{\theta} \|^2 \} = \sigma^2 \, \text{trace}((X^T X)^{-1})$$

Sol.:

The Linear Model opt. solution:

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

$y = X\theta^* + \epsilon$, Substitution

$$\hat{\theta} = (X^T X)^{-1} X^T (X\theta^* + \epsilon)$$

$$\therefore \hat{\theta} = \theta^* + (X^T X)^{-1} X^T \epsilon \rightarrow \theta^* - \hat{\theta} = -(X^T X)^{-1} X^T \epsilon$$

Now: MSE:

$$MSE(\theta^*, \hat{\theta}) = E_\epsilon \{ \| \theta^* - \hat{\theta} \|^2 \}$$

$$= E_\epsilon \{ \| (X^T X)^{-1} X^T \epsilon \|^2 \}$$

Note: ϵ follows Gaussian:

$$\boxed{E(\epsilon \epsilon^T) = \sigma^2 I}$$

$$= Tr(\underbrace{(X^T X)^{-1} X^T}_{I} \underbrace{\epsilon \, \epsilon^T}_{\sigma^2} X (X^T X)^{-1}$$

$$= \sigma^2 \, tr((X^T X)^{-1}) \quad \#$$

Section 3 :- As N increases, MSE decreases.

Sol. :- Woodbury Matrix Identity :- Matrices :- A, U, C and V

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

where:

$$A \in R^{d \times d} \quad A = \sigma^2 I$$
$$U \in R^{d \times N} \quad U = X$$
$$C \in R^{N \times N} \quad C = I$$
$$V \in R^{N \times d} \quad V = X^T$$

So, by N increasing → The rank of X still the same is d. But, covariance improves $(X^T X)$

and as $X^T X$ increases the trace decreases.

$\text{trace}((X^T X)^{-1})$ → Decrease with N increases.

Hence, by Using Woodbury Identity we see:

$\text{trace}((X^T X)^{-1}) \to 0$ as $N \to \infty$

$\therefore$ MSE $\to \sigma^2 \cdot 0 = 0$ by N increases.