LUMS
A Not-for-Profit University

Department of Electrical Engineering
School of Science and Engineering

EE212 Mathematical Foundations for Machine Learning and Data Science

TUTORIAL 2

Tutorial 2-1

Let U be an orthogonal $n \times n$ matrix. Show that its transpose U^T is also orthogonal.

Solution.

If U is square and $U^T U = I$, then U^T is the inverse of U and $UU^T = I$. Since U is the transpose of U^T , this shows that the matrix U^T is orthogonal.

Tutorial 2-2

For an $m \times n$ matrix A and its pseudo-inverse $A^\dagger$, show that $A = AA^\dagger A$ and $A^\dagger = A^\dagger AA^\dagger$ in each of the following cases.

(a) [5 marks] A is tall with linearly independent columns.

(b) [5 marks] A is wide with linearly independent rows.

(c) [5 marks] A is square and invertible.

Solution.

(a) For a tall matrix A with linearly independent columns, $A^\dagger$ is a left inverse of A, so $A^\dagger A = I$. Therefore,

$$AA^\dagger A = A(A^\dagger A) = AI = A, \qquad A^\dagger AA^\dagger = (A^\dagger A)A^\dagger = IA^\dagger = A^\dagger.$$

(b) For a wide matrix A with linearly independent rows, $A^\dagger$ is a right inverse of A, so $AA^\dagger = I$. Therefore,

$$AA^\dagger A = (AA^\dagger)A = IA = A, \qquad A^\dagger AA^\dagger = A^\dagger(AA^\dagger) = A^\dagger I = A^\dagger.$$

(c) For a square invertible A, $A^\dagger = A^{-1}$, so

$$AA^\dagger A = AA^{-1}A = AI = A, \qquad A^\dagger AA^\dagger = A^{-1}AA^{-1} = IA^{-1} = A^{-1} = A^\dagger.$$

Tutorial 2-3

In least squares, the objective (to be minimized) is

$$\|Ax - b\|^2 = \sum_{i=1}^{m} (\tilde{a}_i^T x - b_i)^2,$$

where $\tilde{a}_i^T$ are the rows of A, and the n-vector x is to chosen. In the *weighted least squares problem*, we minimize the objective

$$\sum_{i=1}^{m} w_i (\tilde{a}_i^T x - b_i)^2,$$

where w_i are given positive weights. The weights allow us to assign different weights to the different components of the residual vector. (The objective of the weighted least squares problem is the square of the weighted norm, $\|Ax - b\|_w^2$)

(a) [8 marks] Show that the weighted least squares objective can be expressed as $\|D(Ax - b)\|^2$ for an appropriate diagonal matrix D. This allows us to solve the weighted least squares problem as a standard least squares problem, by minimizing $\|Bx - d\|^2$,where $B = DA$ and $d = Db$.

(b) [7 marks] Show that when A has linearly independent columns, so does the matrix B.

(c) [5 marks] The least squares approximate solution is given by $\hat{x} = (A^T A)^{-1} A^T b$. Give a similar formula for the solution of the weighted least squares problem. You might want to use the matrix $W = \mathbf{diag}(w)$ in your formula.

Solution.

(a) Since the weights are positive, we can write the objective as

$$\sum_{i=1}^{m} w_i (\tilde{a}_i^T x - b_i)^2 = \sum_{i=1}^{m} (\sqrt{w_i} (\tilde{a}_i^T x - b_i))^2 = \|D(Ax - b)\|^2,$$

where D is the diagonal matrix

$$\begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{w_m} \end{bmatrix}$$

(b) We show that $Bx = 0$ implies $x = 0$.
Suppose $Bx = DAx = 0$. Then $Ax = 0$ because D is a diagonal matrix with positive diagonal entries, and hence invertible. By assumption, A has linearly independent columns, so $Ax = 0$ implies $x = 0$.

(c) The solution of the weighted least squares problem is

$$\begin{aligned} (B^T B)^{-1} B^T d &= ((DA)^T (DA))^{-}1 (DA)^T (Db) \\ &= (A^T D^2 A)^{-1} (A^T D^2 b) \\ &= (A^T W A)^{-}1 (A^T W b) \end{aligned}$$

where $W = D^2 = \mathbf{diag}(w)$

Tutorial 2-4

The sum of the diagonal entries of a square matrix is called the trace of the matrix, denoted $\mathbf{tr}(A)$.

(a) [8 marks] Suppose A and B are $m \times n$ matrices. Show that

$$\mathbf{tr}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}$$

What is the complexity of calculating $\mathbf{tr}(A^T B)$?

(b) [7 marks] The number $\mathbf{tr}(A^T B)$ is sometimes referred to as the inner product of the matrices A and B. (This allows us to extend concepts like angle to matrices.) Show that $\mathbf{tr}(A^T B) = \mathbf{tr}(B^T A)$.

(c) [4 marks] Show that $\mathbf{tr}(A^T A) = \|A\|^2$. In other words, the square of the norm of a matrix is the trace of its Gram matrix.

(d) [4 marks] Show that $\mathbf{tr}(A^T B) = \mathbf{tr}(B A^T)$, even though in general $A^T B$ and $B A^T$ can have different dimensions, and even when they have the same dimensions, they need not be equal.

Solution.

(a) The diagonal entries of $A^T B$ are

$$(A^T B)_{jj} = \sum_{i=1}^{m} (A^T)_{ji} B_{ij} = \sum_{i=1}^{m} A_{ij} B_{ij}$$

The complexity is $2mn$ flops. We need mn multiplications and mn - 1 additions. Note that this is lower than the $2mn^2$ complexity of the entire matrix-matrix product $A^T B$.

(b) The matrix $B^T A$ is the transpose of $A^T B$, so it has the same diagonal entries and trace.

(c) From part (a), $A^T A = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 = \|A\|^2$.

(d) We have

$$\mathbf{tr}(B A^T) = \sum_{i=1}^{m} (B A^T)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} B_{ij} (A^T)_{ji} = \sum_{i=1}^{m} \sum_{j=1}^{n} B_{ij} A_{ij}$$

the same expression as in part (a)