Linear Algebra for DS

2024-01-29

- m denotes the number of equations
- n denotes the number of variables

Three cases are possible with m and n

1. m < n, usually multiple solutions are possible
2. m = n, only one possible solution
3. m > n, infinite many solutions are possible.

If all the **rows** of a matrix are **linearly independent**, then **RANK** of the matrix is **FULL** row rank.

Generally, **rows** of the matrix are considered as **data samples**. So, data samples are independent to each other in this case.

CASE:1 $M = N$

Matrix equations with $m = n$

1. Rank of A is full. $rank(A) = n$

- Here, the linear System of equations are consistent.
- If the $det(A)$ is non-zero, then A has unique solution
- $x = A^{-1} * b$

2. Rank of A is not full i.e., $rank(A) < n$

- Here, if the system of equations are consistent, then
  - A has ∞ number of solutions.
- if the system of equations are non-consistent, then
  - A has no solutions.

Given the system of equations:

$$x_1 + 2x_2 = 5$$
$$2x_1 + 4x_2 = 10$$

in matrix form $A * x = b$ as follows:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$

Here, $det(A) = 0 \Rightarrow rank(A) = 1$ and $nullity = 1$

Also, the system of equations are consistent and there exists infinite number of solutions for the given set of equations.

CASE-2: M $>$ N

Here, the number of equations are greater than the number of variables. This is sometimes termed as a no-solution case.

- However, we can identify an appropriate solution by viewing the case from optimization perspective.
- Here, instead of identifying the optimal solution to $Ax - b = 0$, we have to find an optimal x so that $Ax - b$ is minimized.
- Denote $(Ax - b) = e(mx1)$, there are m equations and m errors.
- so, we have to minimize all the errors by minimizing: $\sum_{i=1}^{m} e_i^2$
- This is same as minimizing $(Ax - b)^\top (Ax - b) \Rightarrow \sum_{i=1}^{m} e_i^2$

$(Ax - b)^T = (Ax)^T - b^T = x^T A^T - b^T$

$= (x^T A^T - b^T)(Ax - b)$

$= x^T A^T Ax - x^T A^T b - b^T Ax + b^T b$

Let's focus on the two terms in the middle, $-x^T A^T b$ and $-b^T Ax$:

- The term $-x^T A^T b$ comes from multiplying $-b^T$ with Ax. However, due to the properties of transposition and the fact that the result is a scalar (real number), we can also write $-x^T A^T b$ as $-b^T Ax$, because the transpose of a scalar is the scalar itself. This is why the term appears as $-b^T Ax$ after transposition.
- The term $-b^T Ax$ is just the direct multiplication of $-b^T$ with Ax.

Now, because these two terms are scalars and transposes of each other, they are equal: $-x^T A^T b = -b^T Ax$

$\Rightarrow (Ax - b)^\top (Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b$

Hence, the optimization problem is:

$= min[(Ax - b)^T (Ax - b)]$

$= min[x^T A^T Ax - 2b^T Ax + b^T b]$

$\therefore$ The solution to the optimization problem is given by differentiating $f(x)$ with respect to x and setting the differential to Zero.

$\frac{df}{dx} = 0$

$\frac{df(x)}{dx} = x^T A^T Ax - 2b^T Ax + b^T b$

After splitting the differential, we get:

1. For the first term $x^T A^T Ax$:

The derivative of this term with respect to x is $2A^T Ax$, assuming A is a symmetric matrix

2. For the second term $-2b^T Ax$:

The derivative with respect to x is $-2A^T b$.

3. For the third term $b^T b$:

   This term is a constant with respect to x, so its derivative is 0.

Putting it all together, the gradient of $f(x)$ with respect to x is:

$\nabla_x f(x) = 2A^T Ax - 2A^T b$

$(A^T A)x = A^T b$

Assuming that all the columns are linearly independent

$x = (A^T A)^{-1} A^T b$