LINEAR ALGEBRA: MATRIX ALGEBRA

LINEAR ALGEBRA > QUADRATIC FORMS

A quadratic form at x of a matrix, is a function defined as:

EXPL87

DEFINITENESS OF QUADRATIC FORMS

- A quadratic form, Q, is said to be positive definite if $Q(x)>0$ for all x(all its eigen values are positive)
- A quadratic form, Q, is said to be negative definite if $Q(x)<0$ for all x(all its eigen values are negative)
- A quadratic form, Q, is said to be indefinite if $Q(x)>0$ or $Q(x)<0$ for all x(Its eigen values are either positive or negative)
- A quadratic form, Q, is said to be positive semi-definite if $Q(x)>=0$ for all x(all its eigen values are positive or equals zero)
- A quadratic form, Q, is said to be negative semi-definite if $Q(x)<=0$ for all x(all its eigen values are negative or equals zero)

TWO SIDED INVERSE

The two sided inverse is the classical inverse we have worked with through out this course.

i.e.

EXPL96

LEFT INVERSE and RIGHT INVERSE

The left inverse of a matrix Am,n exists if the columns of A are linearly independent (Hence Col(A) is all the columns of A and Nul(A) = zero vector). Hence (ATA)-1 exists.

Obviously m>n.

The right inverse of a matrix Am,n exists if the columns of AT are linearly independent (Hence Row(A) is made of all the columns of AT and Nul(AT) is the zero vector)

The left and right inverses are defined as:

EXPL97

PSEUDO INVERSE

Now let's consider a matrix Am,n of rank r<m and r<n i.e. not all the columns of A are linearly independent and not all the rows are linearly independent.

Hence the null space and left null space aren't empty.

The inverse of such matrices is known as the pseudo inverse.

We make use of the SVD of the matrix Am,n to define its pseudo inverse.

EXPL98

$$A = \begin{bmatrix} \boxed{\begin{matrix} 3 \\ 3 \\ 3 \end{matrix}} & \begin{matrix} 3 \\ 3 \\ 3 \end{matrix} & \begin{matrix} 3 \\ 3 \\ 3 \end{matrix} \end{bmatrix}$$

- Python Setup, Basic Python, numpy, pandas and matplotlib
- Matrix Algebra
- Linear equations and transformations
- Vectors
- Vector Spaces
- Metric Spaces, Normed spaces, Inner Product Spaces
- Orthogonality
- Determinant and Trace Operator
- Matrix Decompositions (Eigen, SVD and Cholesky)
- Symmetric matrices and Quadratic Forms
- Left Inverse, Right Inverse, Pseudo Inverse

Connecting the dots 1
LINEAR REGRESSION

Connecting the dots 2
PRINCIPAL COMPONENT ANALYSIS

- $b \in Col(A) \dashrightarrow \exists x \,/\, b = Ax$
- $y \in V, H \sqsubseteq V \quad B = Basis(H) \quad min(dist(y, H)) = Proj_H(y)$
- $Proj_H(y) = B(B^T B)^{-1} B^T y$

Matrix Algebra | Metric Spaces, Normed spaces, Inner Product Spaces

Linear equations and transformations | Orthogonality | Vector Spaces

Vectors | Symmetric matrices and Quadratic Forms

COMPLETE LINEAR ALGEBRA

$Ax = b$

$b = Ax \longrightarrow b \in Col(A)$ $\quad \nexists x \,/ b = Ax \longrightarrow b \notin Col(A)$

$\|Ax - b\| \longrightarrow 0$ $\quad \left(\|x\| = \sqrt{\sum x_i^2} \right)$

$\exists\, x', b' \,/ \|Ax' - b'\| = 0$

$\longrightarrow b' \in Col(A)$

$\forall b' \in Col(A), \quad \exists b'_{min}, d(b'_{min} - b) \leq d(b' - b)$

$b'_{min} = Proj_{Col(A)}(b) \quad Col(A) = Span([A_{:,1}, \dots, A_{:,n}])$

$[A_{:,1}, \dots, A_{:,n}] = Base(Col(A))$

$\longrightarrow b'_{min} = A(A^TA)^{-1}A^Tb \longrightarrow Ax'_{min} = A(A^TA)^{-1}A^Tb$

$\longrightarrow A^{-1}Ax'_{min} = A^{-1}A(A^TA)^{-1}A^Tb$

$\longrightarrow x'_{min} = (A^TA)^{-1}A^Tb \longrightarrow [b'_{min}]_E = A(A^TA)^{-1}A^Tb$

$\& \; [b'_{min}]_A = x'_{min} = (A^TA)^{-1}A^Tb$

$A = \begin{bmatrix} 4 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ Try finding a solution for $Ax = b$. If not possible, find its least square (approximate) solution

$$Ax = b \longrightarrow \begin{bmatrix} 4 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \longrightarrow \begin{cases} 4x_1 + 1x_2 = 1 \\ 0x_1 + x_2 = 3 \\ 1x_1 + 1x_2 = 1 \end{cases}$$

$$\left[\begin{array}{cc|c} 4 & 1 & 1 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{array}\right]\begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array} \sim \left[\begin{array}{cc|c} 4 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & \frac{3}{4} & \frac{3}{4} \end{array}\right]\begin{array}{l} r_1 \\ r_2 \\ r_3 = r_3 - \frac{1}{4}r_1 \end{array}$$

$$\sim \left[\begin{array}{cc|c} 4 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 1 & 1 \end{array}\right]\begin{array}{l} r_1 \\ r_2 \\ r_3 = \frac{4}{3}r_3 \end{array} \sim \left[\begin{array}{cc|c} 4 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{array}\right]\begin{array}{l} r_1 \\ r_2 \\ r_3 = r_3 - r_2 \end{array} \quad 0 \neq -2$$

$$x'_{min} = (A^TA)^{-1}A^Tb \quad A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 5 \\ 5 & 3 \end{bmatrix}$$

$$det(A^TA) = (17 * 3) - (5 * 5) = 26$$

$$\longrightarrow (A^TA)^{-1} = \frac{1}{26}\begin{bmatrix} 3 & -5 \\ -5 & 17 \end{bmatrix}$$

$$(A^TA)^{-1}A^T = \frac{1}{26}\begin{bmatrix} 3 & -5 \\ -5 & 17 \end{bmatrix}\begin{bmatrix} 4 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{26}\begin{bmatrix} 7 & -5 & -2 \\ -3 & 17 & 12 \end{bmatrix}$$

$$(A^TA)^{-1}A^Tb = \frac{1}{26}\begin{bmatrix} 7 & -5 & -2 \\ -3 & 17 & 12 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = \frac{1}{26}\begin{bmatrix} -10 \\ 60 \end{bmatrix}$$

$$\boxed{\longrightarrow x'_{min} = \frac{1}{26}\begin{bmatrix} -10 \\ 60 \end{bmatrix}}$$

$$Ax'_{min} = b'_{min} = \frac{1}{26}\begin{bmatrix} 4 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -10 \\ 60 \end{bmatrix} = \frac{1}{26}\begin{bmatrix} 20 \\ 60 \\ 50 \end{bmatrix}$$

$$\rightarrow \forall b' \in Col(A), \|b - b'_{min}\| \leq \|b - b'\|$$

$$\longrightarrow \left\| \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.77 \\ 2.31 \\ 1.92 \end{bmatrix} \right\| \leq \left\| \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - b' \right\|$$

In Class Exercise 1

Solve $Ax = b$ and also find its least square solution

$A = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

In Class Exercise 1 Solution

$$Ax = b \longrightarrow \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \longrightarrow \begin{cases} 2x_1 + 0x_2 = 2 \\ 1x_1 + 3x_2 = 4 \end{cases}$$

$$\left[\begin{array}{cc|c} 2 & 0 & 2 \\ 1 & 3 & 4 \end{array}\right]\begin{array}{l} r_1 \\ r_2 \end{array} \sim \left[\begin{array}{cc|c} 2 & 0 & 2 \\ 0 & 6 & 6 \end{array}\right]\begin{array}{l} r_1 \\ r_2 = 2r_2 - r_1 \end{array}$$

$$\longrightarrow \begin{cases} 2x_1 = 2 \\ 6x_2 = 6 \end{cases} \longrightarrow \begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$$

$$x = (A^TA)^{-1}A^Tb = \left(\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}\right)^{-1}\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

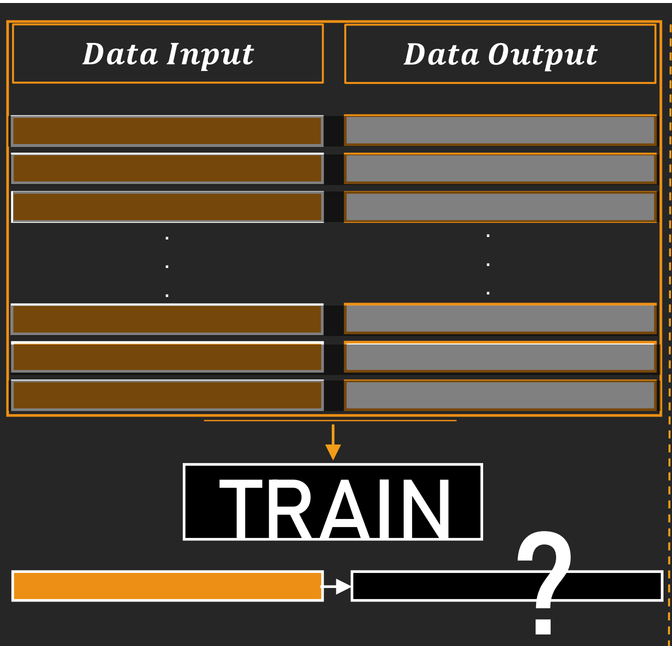
$$= \left(\begin{bmatrix} 5 & 3 \\ 3 & 9 \end{bmatrix}\right)^{-1}\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$= \frac{1}{36}\begin{bmatrix} 9 & -3 \\ -3 & 5 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$= \frac{1}{36}\begin{bmatrix} 18 & 0 \\ -6 & 12 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \frac{1}{36}\begin{bmatrix} 36 \\ 36 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

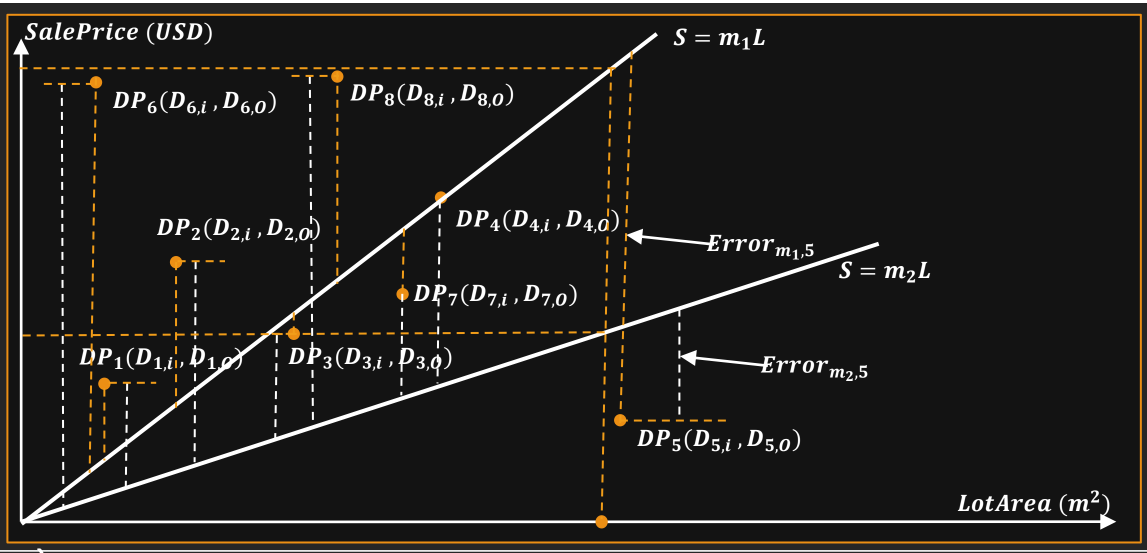
$$b = Ax = A(A^TA)^{-1}A^Tb = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

$$Col(A) = Span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}\right\} \qquad \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 1 * \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 * \begin{bmatrix} 0 \\ 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 \\ 4 \end{bmatrix} \in Col(A)$$



LotArea	*SalePrice*
$D_{1,i}$	$D_{1,0}$
$D_{2,i}$	$D_{2,0}$
$D_{3,i}$	$D_{3,0}$
$D_{4,i}$	$D_{4,0}$
$D_{5,i}$	$D_{5,0}$
$D_{6,i}$	$D_{6,0}$
$D_{7,i}$	$D_{7,0}$
$D_{8,i}$	$D_{8,0}$

$Error \approx 0$



$$\longrightarrow Best\ Error = \min\sqrt{\left((Error_{m_i,1})^2 + (Error_{m_i,2})^2 + (Error_{m_i,3})^2 + (Error_{m_i,4})^2 + (Error_{m_i,5})^2 + (Error_{m_i,6})^2 + (Error_{m_i,7})^2 + (Error_{m_i,1})^2\right)}$$

$$= \min\sqrt{\sum_{j=1}^{8}\left((Error_{m_i,j})^2\right)} = \min\sqrt{\sum_{j=1}^{8}\left((mL_j - S_j)^2\right)} = \min\|Lm - S\| \leftrightarrow \min(\|Ax - b\|)\quad A = L,\quad X = m,\quad b = S$$

$$x'_{min} = (A^TA)^{-1}A^Tb \longrightarrow m'_{min} = (L^TL)^{-1}L^TS = \left(\begin{bmatrix} D_{1,i} & \cdots & D_{8,i}\end{bmatrix}\begin{bmatrix} D_{1,i} \\ \vdots \\ D_{8,i}\end{bmatrix}\right)^{-1}\begin{bmatrix} D_{1,i} & \cdots & D_{8,i}\end{bmatrix}\begin{bmatrix} D_{1,0} \\ \vdots \\ D_{8,0}\end{bmatrix} \in R^{1x1}$$

$D_{9,i}$???

$$\boxed{D_{9,0} = m'_{min}D_{9,i}}$$

	Input 1	*Input 2*	*Input 3*	*Output*
DP 1	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,0}$
DP 2	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$	$D_{2,0}$
DP 3	$D_{3,1}$	$D_{3,2}$	$D_{3,3}$	$D_{3,0}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
DP 1460	$D_{1460,1}$	$D_{1460,2}$	$D_{1460,3}$	$D_{1460,0}$

Input 1 = MSSubClass(0 → 320) Input 2 = LotFrontage(m^2)

Input 3 = LotArea(m^2) Output = SalePrice(USD)

$$Input = I = \begin{bmatrix} D_{1,1} & D_{1,2} & D_{1,3} \\ \vdots & \vdots & \vdots \\ D_{1460,1} & D_{1460,2} & D_{1460,3}\end{bmatrix} \quad Output = O = \begin{bmatrix} D_{1,0} \\ \vdots \\ D_{1460,0}\end{bmatrix}$$

$O = I(m), has\ no\ solution$

$$m'_{min} = m'_{optimal} = (I^TI)^{-1}I^TO = \begin{bmatrix} m_1 \\ m_2 \\ m_3\end{bmatrix} \in R^{3x1}$$

$$New\ Input = \begin{bmatrix} D_{1461,1} & D_{1461,2} & D_{1461,3}\end{bmatrix} \quad New\ Output = New\ Input(m'_{optimal}) = \begin{bmatrix} D_{1461,1} & D_{1461,2} & D_{1461,3}\end{bmatrix}\begin{bmatrix} m_1 \\ m_2 \\ m_3\end{bmatrix}$$

$$Input = I = \begin{bmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m,1} & D_{m,2} & \cdots & D_{m,n}\end{bmatrix} \quad Output = O = \begin{bmatrix} D_{1,0} \\ \vdots \\ D_{m,0}\end{bmatrix} \rightarrow m'_{optimal} = (I^TI)^{-1}I^TO = \begin{bmatrix} m_1 \\ \vdots \\ m_n\end{bmatrix} \in R^{nx1}$$

$$New\ Input = \begin{bmatrix} D_{m+1,1} & \cdots & D_{m+1,n}\end{bmatrix} \quad New\ Output = New\ Input(m'_{optimal}) = \begin{bmatrix} D_{m+1,1} & \cdots & D_{m+1,n}\end{bmatrix}\begin{bmatrix} m_1 \\ \vdots \\ m_n\end{bmatrix}$$

LINEAR ALGEBRA > LEAST SQUARES AND APPLICATION TO LINEAR REGRESSION

Under linear equations and transformations, we saw that an equation Ax = b, can have no solution at times. It should be noted that, for b = Ax, b should be part of the col space of A.
So when Ax=b has no solution, it implies that b isn't part of the col space of A.

In order to solve such problems, we come up with a least square solution which minimizes the difference between b and Ax. Such that
EXPL89
In a previous section, we saw that the shortest distance from a point to a subspace is the orthogonal projection from that point to the subspace.
We can find b' which is the orthogonal projection of b on the col space of A (In order to minimize the distance between b and b'). with b' = Ax'.
If the columns of A are linearly independent, then The column space is made of all the column vectors of A. AND ATA is invertible.
EXPL90

We can make apply our knowledge on least squares in a popular machine learning algorithm which is linear regression.
According to Andrew Ng (Adjunct Professor at Stanford University), Machine learning is the science of getting computers to act without being explicitly programmed.
In order for them to act without being explicitly programmed, they rely on much data, sometimes soo much data.

LINEAR ALGEBRA > LEAST SQUARES AND APPLICATION TO LINEAR REGRESSION

How Machine Learning algorithms work is that,

EXPL91

We are given several data inputs and their corresponding output(s). This is called the dataset.
We make the algorithm learn from this input- output pairs and then later on, when given an input, we can predict the output.
Let's take an example from a part of the Kaggle House price competition.
Kaggle is a datascience platform where users can share, collaborate and compete. Other very popular datascience platforms include zindi, drivendata and datahack.

EXPL92

In this example, our dataset is made up of 1460 different data points, with each data point made of 3 inputs and one output.
The inputs are the characteristics of a house to be sold and then the output is the sale price of the house.
The first input is the building class (MSSubClass), which is the quality of the buildings.
The second input is the building frontage (LotFrontage), which describes the amount of space left infront of the building, just before the next building.
The third input is the building size (LotArea), which is describes the amount of space occupied by the building.

Then the output, which is the sale price of the building.

LINEAR ALGEBRA > LEAST SQUARES AND APPLICATION TO LINEAR REGRESSION

After collecting and cleaning all this data, several machine learning algorithms can be used to infer on the sale price of a building, depending on these three factors.

A very popular Machine learning algorithm which can be used to solve this problem is known as linear regression.
Let's see how it works.
EXPL93
Consider only one of the inputs (Lot Area), it is directly related to the sale price in most cases, although we may come across some outliers.
The linear regression algorithm seeks to find the best fit line, which permits to characterize the dataset.
So learning from this dataset entails finding the best fit line.
To get the best fit line, we try to minimize the difference between each point and the best fit line.
Going back to our original problem with three inputs,
EXPL94
In Machine learning literature, m is called the weights. So you'll generally hear Machine Learning practitioners talk of learning from a dataset, in order to obtain weights.
And then use these weights for inference.
Which is just what we just did.

Then complete practicals for all the topics
Also complete exercises for all the topics

LINEAR ALGEBRA: MATRIX ALGEBRA

RATIONALE

SDFDF
ASDFDFSDF
ASDDFDFSAFDSF