Matrix Analysis with Applications
Dr. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 27
Least Square Approximation

Hello friends. So, welcome to the lecture on least square approximation. So, in this lecture we will learn how to find out an approximate solution to a linear system, which is either over-determined and not having the exact solution and in other case if it is underdetermined and having the infinitely many solutions. So, out of those infinitely many solution which of the solution will be having the minimum error.

(Refer Slide Time: 01:01)

Consider a linear system $AX = b$, where matrix

(I) A is square matrix of order n. Moreover, assume that A is a full rank matrix (A is nonsingular)

$\boxed{X = A^{-1} b}$ exact and unique

(II) If A is a $m \times n$ matrix, where $m >> n$
(means, more equations than unknowns)
In this case, the system is an overdetermined system. Moreover, assume that the system does not have an exact solution. Hence, we need to find an approximate solution which minimize the residual error.

$$E = \|AX - b\|^2$$

$$\min_X E \Rightarrow E = (AX-b)^T(AX-b) = X^TA^TAX - X^TA^Tb - b^TAX + b^Tb$$
$$= X^TA^TAX - 2X^TA^Tb + b^Tb$$

$$\frac{\partial E}{\partial X} = 0 \Rightarrow 2A^TAX - 2A^Tb = 0$$
$$\Rightarrow X = \underline{(A^TA)^{-1}A^T}\, b$$

Consider a linear system A X equals to b, where matrix A is square matrix of order n. So, it means what we are having we are having n equations in n unknown. Moreover, assume that A is a full rank matrix, means the rank of A is n or in other one A is nonsingular.

So, since A is nonsingular so, A inverse exist and the solution of this system can be written as X equals to A inverse into b and this solution will be exact and unique solution to this linear system. Now, this is my first case now second case, if A is not square if A is a m by n matrix means the coefficient matrix is having a rectangular shape where m is greater than n.

So m is greater than n means you are having the more number of equations, then unknown means more equations than unknowns, in this case the system is an over-determined linear system. So, over determined means you are having more equations than unknown. Moreover, assume that the system does not have an exact solution if the system is not having any exact solution then we need to find out an approximate solution.

Hence so, I need to find out an approximate solution and this approximate solution minimize the residual error means the solution with minimum residual error. So, let me define the residual error is e equals to Euclidean norm of A x minus b. So, what I need to do I need to minimize E. So, I have to find out such x, which minimize E ok. So, this means I need to minimize A x minus b transpose into A x minus b, which is the this norm of E, which is nothing just E.

So, since I have to find out minimization of this function. So, what I need to do I have to differentiate E partially with respect to X and I need to put it equals to 0. So, basically from here this E can be written as X transpose A transpose A into X minus X transpose A transpose into b, minus b transpose A X plus b transpose into b. And, this will be equals to X transpose A transpose A X minus 2 X transpose A transpose b plus b transpose b, because these 2 things will be same.

Now, del E over del X equals to 0 will give you twice of A transpose into A X minus when I will do for this one twice of A transpose b and this will be 0. So, this equals to 0. So, what I am getting from here I will get. So, I will get X equals to A transpose A inverse into A transpose into b.

(Refer Slide Time: 08:27)

If we write $(A^TA)^{-1}A^T = A^+$

Least square soln $\Rightarrow$ $X = A^+ b$

Here, $A^+ (= (A^TA)^{-1}A^T)$ is called the Pseudo-inverse of A.

The solution $X = A^+ b$ is called least square approximation of the over determined system $AX = b$.

If we take, A as square and invertible matrix.

$$AA^+ = A[(A^TA)^{-1}A^T] = \underline{A[A^{-1}}\underline{(A^T)^{-1}A^T]}$$
$$= I \cdot I = I$$

Remark:- $A_{m \times n}$ $(m > n)$

A^TA should be invertible

$\text{rank}(A) = \underline{\underline{n}}$

Please note it ~~that~~

So, if we write A transpose A inverse into A transpose as A plus, then least square solution is given as X equals to A plus b where this A plus is A transpose A inverse into A transpose.

Here, A plus which is equals to again I am writing this is called the Pseudo inverse of A. The solution X equals to A plus b that is pseudo inverse of A into b is called least square approximation of the over determined system A X equals to b or least square solution of the over-determined system A x equals to b.

Why I am saying it A pseudo inverse of A, because if you take. So, if I take A as square and invertible matrix means as a full rank matrix then A into A plus will become A into A transpose A inverse into A transpose. So, this will become A into A inverse into because A transpose A inverse will become A inverse into A transpose inverse into A transpose.

So, it means this will be I this will be I because A is A square and invertible. So, A transpose will be invertible. So, this is I into I equals to I. So, it is a kind of inverse only, but in case of rectangular matrices and where number of rows are more than number of columns. So, this is the least square approximation of an over-determined system.

So, let us take an example of it. Here, I am having a remark we have assumed that A is m by n matrix where m is greater than n. So, and for calculating the pseudo inverse of A I am using the inverse of A transpose into A. So, here A transpose A should be invertible and what will be the size of A transpose A it will be of n by n matrix.

So, it means the rank of A should be n then only you can calculate A transpose A inverse and you can apply this pseudo inverse and you can find out the least square approximation in this case. So, please note that. So, please note it if it the rank of A is less than n I will tell you how to find out the least square approximation of the system X equals to b that in the subsequent lectures.

(Refer Slide Time: 13:38)

Example:- Fit the best line from the data points $\underset{\mathrm{I}}{\underline{(1,2)}}$, $\underset{\mathrm{II}}{\underline{(2,3)}}$ and $\underset{\mathrm{III}}{\underline{(3,5)}}$

Solution Assume that line is given as $y = mx + c$

$$\left.\begin{array}{l} \mathrm{I} \Rightarrow 2 = m + c \\ \mathrm{II} \Rightarrow 3 = 2m + c \\ \mathrm{III} \Rightarrow 5 = 3m + c \end{array}\right] \quad \underset{A}{\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}} \underset{X}{\begin{bmatrix} m \\ c \end{bmatrix}} = \underset{b}{\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}} \quad \left(\begin{array}{l} \text{no. of eq}^n = 3 \\ \text{no. of var.} = 2 \end{array}\right)$$

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}; \quad A^T A = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} \frac{1}{2} & -1 \\ -1 & \frac{7}{3} \end{bmatrix}$$

$$X = (A^T A)^{-1} A^T b = \begin{bmatrix} \frac{1}{2} & -1 \\ -1 & \frac{7}{3} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{3} \end{bmatrix}$$

$m = \frac{3}{2}$ and $c = \frac{1}{3}$

$$\boxed{y = \frac{3}{2}x + \frac{1}{3}c} \qquad E = \|y - mx - c\|^2 = \underline{0.1667}$$

Now, let me take an example of it. So, fit the best line from the data points 1 2 2 3 and 3 5. So, we are having 3 points and we have to find out we have to fit the best line from these data. So, let us solve this particular example.

So, here assume that line is given as y equals to m x plus c. So, from the first data point I can write y is 2 2 equals to x is 1 m plus c. Similarly, from the second data point I can write 3 equals to 2 m plus c. And, from the third data point I can write 5 equals to 3 m plus c. So, this particular system can be written as 1 1 2 1 3 1 multiplied with m c and this equals to 2 3 5.

So, this is my matrix A this is X this is b. So, it is A linear system which is over-determined because I am having 3 equations and my unknowns are 2 means to fit the line I have to find out the value of m and c.

So, here if I write the number of equations is 3 and number of variables is 2. So, it is an over-determined system. So, now, use the least square approximation technique. So, first I will calculate A transpose. So, A transpose will come 1 1 2 1 and 3 1.

So, once I calculate A transpose, then I will calculate A transpose into A which will be A 2 by 2 matrix. So, here if I calculate A transpose into A. So, 1 into 2 2 4 3 into 9 plus 4 into 1 it comes out to be 14, then this comes out to be 6 then this comes out to be 3 plus 3 6 and then 3. So, this is A transpose into A.

Now, if the rank of this is 2 then I can calculate the inverse of this. So, here I am calculating the inverse because it is A full rank matrix full rank means rank is 2. So, inverse comes out to be 1 by 2 minus 1 minus 1 and 7 by 3.

Now, least square solution of this system can be given as X equals to A T A inverse into A transpose into b. So, A transpose A inverse is 1 by 2 minus 1 minus 1 7 by 3 into A transpose. So, A transpose will be 1 1 2 1 3 1 into b. So, here b is 2 3 5 if I multiply these 3 matrices it comes out to be 3 by 2 and 1 by 3.

So, it means m is 3 by 2 and c is 1 by 3. So, here best line is y equals to 3 by 2 X plus 1 by 3 C, which is going from these data points. Here the residual error is which is just y minus m x minus c whole square and it is 0.1 6 6 7 which is the minimum 1. So, hence this is the overall idea of applying least square approximation for the systems those are over-determined and here we have taken the case of line fitting.

Now, consider the third case.

(Refer Slide Time: 19:49)

(III) Consider a system $Ax = b$, where A is a $m \times n$ matrix and $m < n$. It is an underdetermined system. This system will be having infinitely many solutions. Now, we need to pick one of these solutions by finding smaller one, i.e.

$$\min \|x\|^2$$
$$\text{subject to } Ax = b$$

By using the method of Lagrange multiplier, we have

$$E = \|x\|^2 + \lambda^T(b - Ax)$$

$$\frac{\partial E}{\partial x} = 0 \Rightarrow 2x - A^T\lambda = 0 \quad \text{— (1)}$$

On premultiplying both sides by A, we have

$$2Ax - AA^T\lambda = 0$$
$$2b - AA^T\lambda = 0 \Rightarrow \lambda = 2(AA^T)^{-1}b$$
$$\Rightarrow 2x - 2A^T((AA^T)^{-1}b) = 0$$
$$\Rightarrow x = A^T(AA^T)^{-1}b = A^+b$$

Consider A system A x equals to b, where A is a m by n matrix and here m is less than n means I am having less equation than the variables. So, it is an underdetermined system. This system will be having infinitely many solutions.

Now, we need to pick one of these solutions by finding smaller one, that is we have to minimize norm of x subject to A X equals to b ok. So, here how to solve it by using the method of Lagrange multiplier, we have our Lagrangian edge E equals to norm of X square plus lambda times t into b minus A X. Now, to minimize this and finding the X I have to write del E by del X equals to 0 which gives 2 X plus A transpose lambda equals to 0 let me write this equation as 1.

Now, I cannot find the value of lambda directly from here, because I am having A transpose in the multiplication with lambda and A is A rectangular matrix. So, the A transpose and I cannot calculate the inverse of A transpose. So, I cannot find out this lambda from here.

So, what I will do on pre-multiplying both sides by A, we have so, I am multiplying by A in both side I will be having 2 A x plus A into A transpose into lambda equals to 0. And sorry it will be minus, because A X with minus sign here and similarly this will be minus. Now, as you know my system is A X equals to b. So, this A X I can replace by b. So, this will become 2 b minus A into A transpose lambda equals to 0.

So, from here I can write lambda equals to A into A transpose inverse into b provided the transpose of A into sorry the inverse of A into A transpose exist. And, it will exist when the matrix A will be having the rank m, if it is less than m then this inverse will not exist because then A into a transpose will become singular.

Now, I am having this one now putting the this value of lambda in equation 1 I will be having 2 X minus A transpose and then lambda is here supporting this value of lambda. So, it will become A into A transpose inverse into b equals to 0 I forgot to write 2 here. So, 2 will come here also.

So, it means X will become A transpose into A A transpose inverse into b. So, if I write this particular matrix as A plus. So, this will become A plus b.

(Refer Slide Time: 26:38)

Here, the matrix $A^+ = A^T(AA^T)^{-1}$ is called the pseudo inverse of A in underdetermined case.

The least square solution is given as

$$X = A^+ b$$

$\longrightarrow$ Overdetermined $\Rightarrow$ $\underline{X = (A^TA)^{-1}A^T b}$

$\longrightarrow$ Undetermined $\Rightarrow$ $\underline{X = A^T(AA^T)^{-1} b}$

$(A^TA)_{n\times n}$ with rank $\leq m$

$\underline{m < n}$

$(AA^T)_{m\times m}$ with rank $\leq \underline{\underline{m}}$

So, here the matrix A plus, which is equals to A T into A into A transpose inverse is called the pseudo inverse of A in underdetermined case. So, please see the difference in over-determined case I was having this matrix as A transpose A inverse into A transpose. Here, I am having A transpose into A A transpose inverse.

So, with this what I can have and that the least square solution is given as X equals to A plus b. So, what we have learnt from here that if system is over-determined, then solution is given as X equals to A transpose into A inverse into A transpose into b.

If, the system is underdetermined means less equation than unknown then the least square solution is given as A transpose A A transpose inverse into b, you must be surprising why I have not used this form in case of underdetermined system. Because, if I will use this form for underdetermined system then my matrix A T A will be of size n by n with rank less than equals to m. So, m will be less than n in underdetermined case.

So, hence this matrix will be always singular and you cannot calculate A transpose A inverse, that is why we have taken this form of the solution. Because, now we have taken A into A transpose which will be A matrix of m by m with rank less than equals to m. So, if rank is equals to m then you can apply this method or least square approximation of this form. If rank is less than m then how to find out least square solution that I will discuss in the next to next lecture. .

So, this is the underdetermined form means the least square solution for the underdetermined system.

(Refer Slide Time: 30:23)

Ex:. Find the LSA of the linear system

$$\left.\begin{aligned} 2x_1 + x_2 + x_3 &= 4 \\ 2x_1 - x_2 + x_3 &= 2 \end{aligned}\right\} \quad \begin{aligned} x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= 1 \end{aligned}$$

Here, $A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & -1 & 1 \end{bmatrix}$

$$AA^T = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix} \Rightarrow (AA^T)^{-1} = \begin{bmatrix} 0.3 & -0.2 \\ -0.2 & 0.3 \end{bmatrix}$$

Now LSA is

$$X = A^T (AA^T)^{-1} b = \begin{bmatrix} 1.2 \\ 1.0 \\ 0.6 \end{bmatrix} \text{ Ans.}$$

Let us take an example of it find the least square approximation of the linear system 2 x 1 plus x 2 plus x 3 equals to 4 and 2 x 1 minus x 2 plus x 3 equals to 2. So, here A is 2 1 1 2 minus 1 1 now A into A transpose will become 6 4 4 6 which is invertible. So, from here I can calculate A into A transpose inverse, which comes out to be 0.3 minus 0.2 minus 0.2 0.3.

And, now least square approximation is X equals to A transpose A into A transpose inverse into b and this comes out to be 1.2 1.0 0.6 means x 1 is 1.2 x 2 is 1 x 3 is 0.6.

One of the solution of this system you can see directly x 1 is 1 x 2 is 1 and x 3 is 1. This solution is satisfying these two equations exactly; however, we are getting a different solution using this least square approximation technique. Why? Because while deriving this method I have told you, I have to look out of the infinite solutions I have to look the solution which is having the minimum norm of x. And the norm of x in this case will be the minimum for a solution which satisfying these two equations.

So, in this lecture we have learnt least square approximation for over-determined as well as underdetermined systems. We have taken couple of example for such type of systems and how to apply least square approximation on those. However, few questions are unanswered here like if A transpose A inverse in case of over-determined system and A into A transpose inverse in case of underdetermined system do not exist, then how to apply these techniques?.

I will speak about it in next to next lecture, because for applying approximation of finding the approximate solution we have to use singular value decomposition tool and that I will discuss in the next lecture, and then how to find out least square approximation using singular value decomposition that will come in the next to next lecture.

Thank you very much.