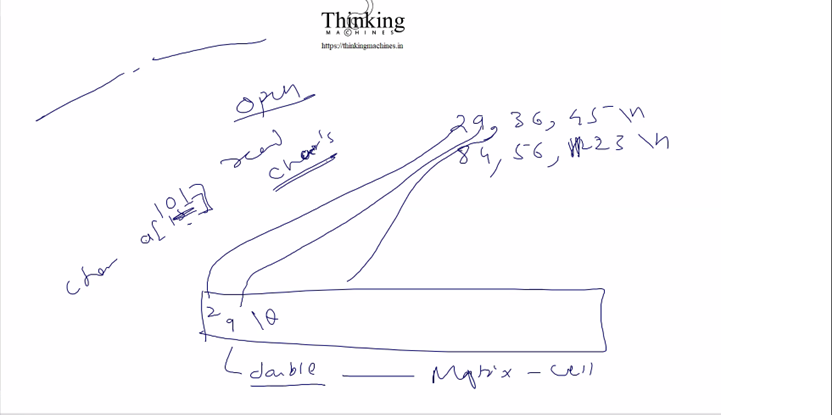
1. **Ab CSV read krne ka code: Comma separated values**

****

Maan lo values ese he 29,36,25 \n 84,56,23\n

**Algorithm:**

Humare matrix me 12 cell he supoose and each cell contains char \* type value

|  |  |  |
| --- | --- | --- |
| 2 | 9 | \0 |

Ab hum har cell k char \* ko ek array assign krege jisme value ese hogi

Let suppose first cell: 2 read krege then 9 read krege and so on , aane pr rukege and\0 likhege.

Then is array ko then double me convert krege.

Is double ki 8 byte me hi fractional and integer ka gyan rkhege he.

Iske liye process of find double ka binary and . Vo saari info 64 bits me aa jayegi.

Now humare matrix me kitne rows and kitne column bnege?

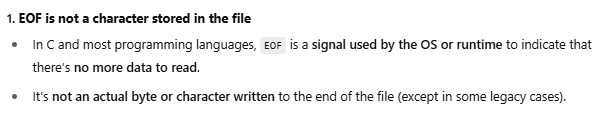
**Ans:**

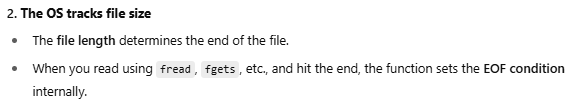
\n se no. of **rows** pta & (no of comma +1) se **columns** ka pta

Note: Check sbme equal no. of comma(,) hona chahiye nahi hua to file ko nahi pdege.

Dono ko ek hi round me pta kr lege.

Before proceeding let’s focus on what is EOF:





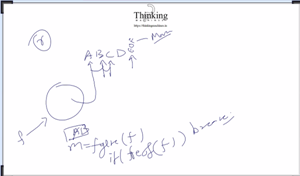
fgetc() as well

fgetc() read one byte in one time, that is one character.

fread() used to read custom bytes,



Kese kaam krta he:



Let suppose ki ek file me ABCD likha he humne FILE \*f=fopen(file)

At first f is pointing to A.

jese hi humne fgetc(f) kiya f returns A and start pointing to B

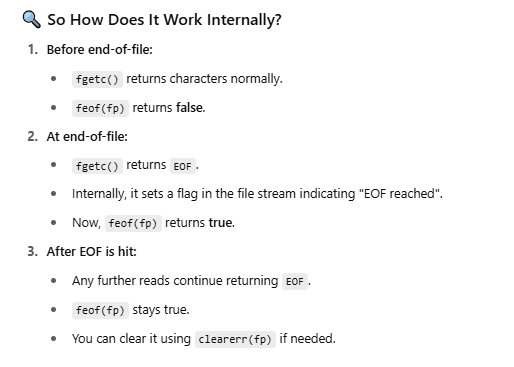
Phir se fgetc(f) kiya f returns B and start pointing to C

Phir se fgetc(f) kiya f returns C and start pointing to D

Phir se fgetc(f) kiya f returns D and start pointing to EOF

Ab is baar if(feof(f)) break me fas jayega, isko ese bhi kr skte he coz vo true return krega

If(fgetc(f)== EOF) break;





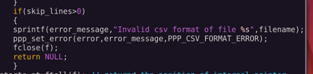
EOF ko bhi as a data treat krega and next loop me break krega

Ye he humara example



Ye code isliye likha he ki user skip\_lines ka argument pass krega . maan lo csv me 5 lines he and skip lines user ne >=5 de di to saari hi line skip ho jayegi and read k liye kuch nahi bchega

To us case k liye ye check lagaya ki saare \n khtm hone k baad bhi skiplines bche to saari line skip ho gayi he



Condition me >0 kyu liya

Let suppose csv me 5 lines and user ne 4 \n diye and last wali line pr \n nahi

Hmara code

Skiplines=5, \n=4

R1: skiplines =4 , \n=3

R2: skiplines =3 , \n=2

R3: skiplines =2 , \n=1

R4: skiplines =1 , \n=0 bche

Iske baad read krte krte EOF aa jayega and loop break

Loop break k baad skiplines >1 to return

Next:

Skiplines=5, \n=5

R1: skiplines =4 , \n=4

R2: skiplines =3 , \n=3

R3: skiplines =2 , \n=2

R4: skiplines =1 , \n=1 bche

R5: skiplines =0 , \n=0 bche

Final: skiplines cuts \n, and maan lo skiplines ko cut krte krte \n hi khtm ho gaye and phir bhi skiplines bchi means hum kya hi skip kre and kya hi read kre.

Means 0 rows remaining to read after skipping all lines.

Ab skip line k baad actual line se kese start kre?

Jb while(skiplines>0) wala loop without feof() me fse break hoga means humne last skipped line ka \n read kiya and now f points to first character of actual line.

1,2,3,\n

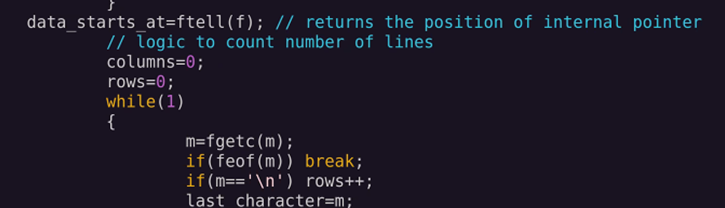
2,2,3\n

3,4,5\n

123,24,344\n

Maan lo skiplines=3 tha to jb loop break hoga tb pointer 123 ko point kr raha hoga kyuki 3rd line ka \n read krne k baad hi skipline—hone se skipline ki value 0 hui hogi and loop break ho gaya hoga.

So, we capture this address into a variable and we will start counting comma and \n from that point.



Ab is point k baad ek baar to traverse krege to get no. of columns[,] and rows[\n]

Then phir se isi point pr aake actual data ko read krege.

Is baar fget() se char read krege binary nahi .

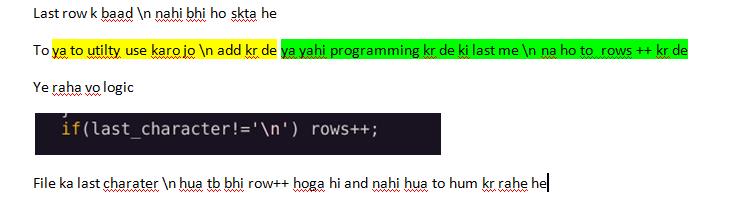
**Algo:**

1. **Capture position after skipping lines , which is start of actual data.**
2. Count no. of lines and columns [no. of comma(,) in each line should equal other wise invalid format error]
3. Create matrix after getting data[rows and columns]
4. Move file pointer back to data\_start\_at using fseeek()
5. Read data from file, one by one byte and populate matrix
6. Close file
7. Return matrix

Iske phle k steps the:

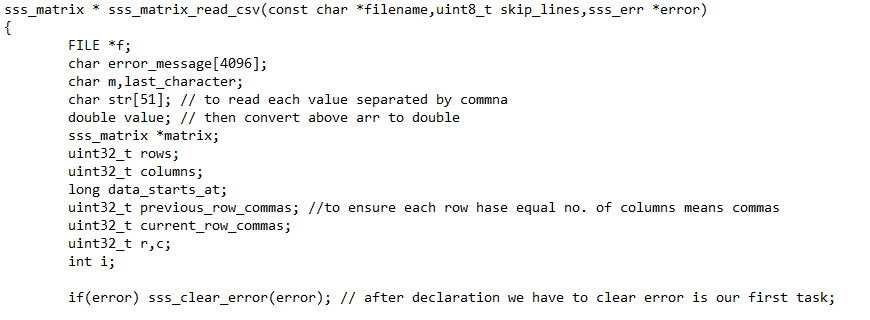
1. Verify matrix and filename is not null
2. Verify file opened successfully
3. Verify no. of lines in file is greater then skiplines.

Is algo me ek issue he ki last line pr \n na bhi mile

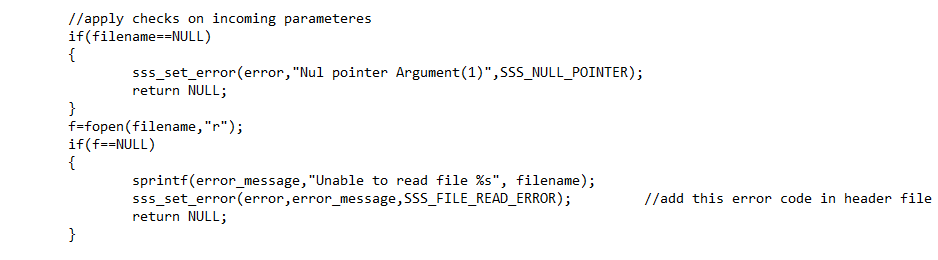


**code of read csv**: in matrix.c file

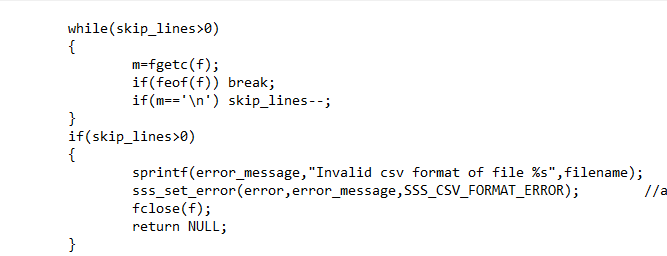
Declaration and clear error:



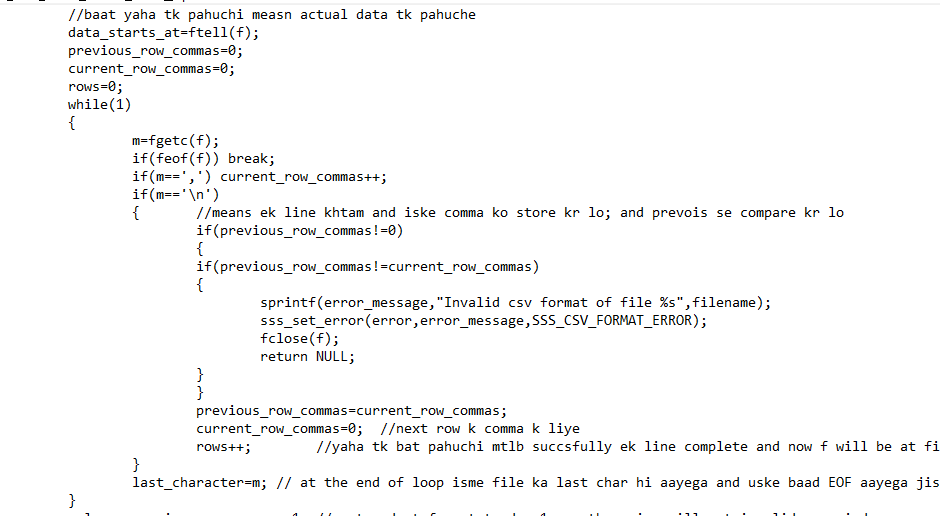
Apply check for filename and fopen()



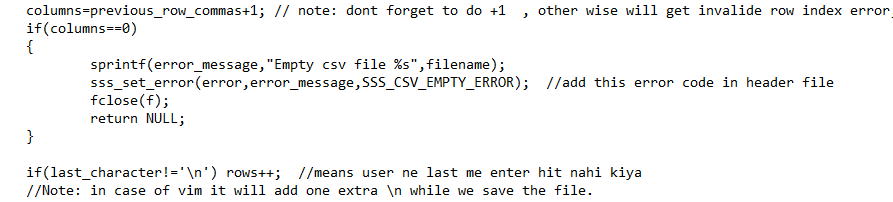
Skip lines:



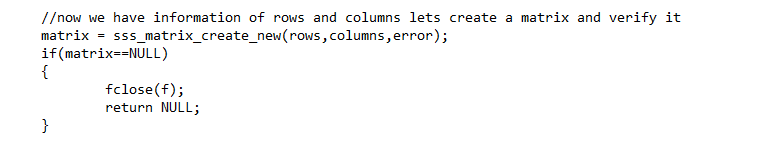
Keep cuurent position in a var and take a loop over file and count comma and \n



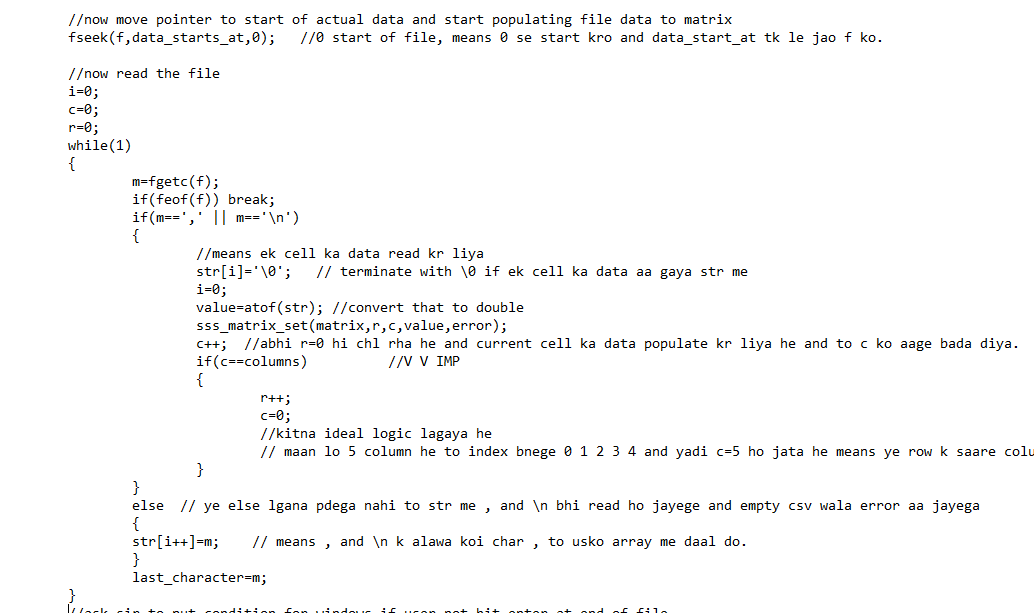
Verify no. of columns and do row++ in case user not put \n or hit enter at end



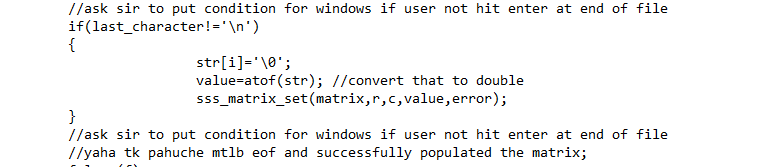
Create matrix:



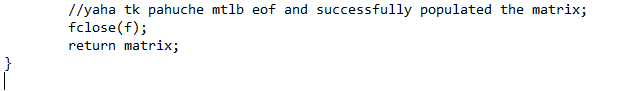
Move pointer to actual data position again and loop to read the actual data now



Now if user not put \n tb last wali value k liye if me nahi ghusega and usko 0 read kr liya jayega default. To us last value k liye ek or condition lgani pdegi, str me jo data pda he usko lekr matrix me set. Below part is intentionally not covered by sir , they will look later.



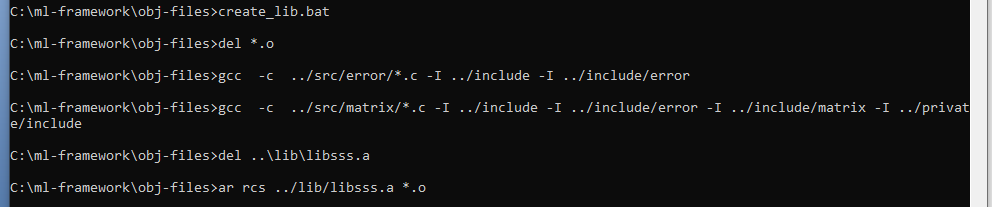
Close file and return matrix



Add error code:



Compile and create lib



Write test case



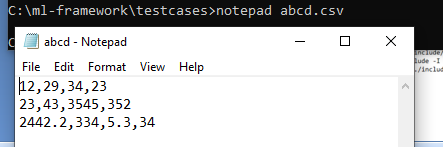


Add line to compile this test case:

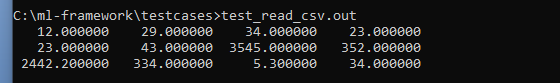


Run batch file to compile and then

Create a csv:



Now run the executable: and below is final output

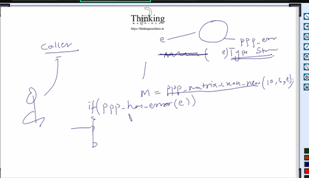


Abhi generic ka idea: think not to implement, implement baad me sir krege.

Data ka type void \* kr de to kya change aayega.

Ek char bdega to chek ki kis type ka , like D for double, I for integer etc

1. **Ek issue he:**

****

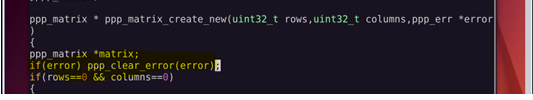
ek line pr matrix create ko call kiya and next pr check kiya ki has\_error()

to kya matrix successfully create hone k baad bhi ye next line ki if me fasega ?

ans: yes , kyuki ho skta he ji line pr create\_matrix call kiya uske phle kisi dusre function ko call kiya ho and vo fail ho gaya ho to usne error ko update kr diya ho .

Solution: in each function jaha bhi error ka object aa raha he] after variable declaration clear error call krna he. Taki pichle function call se koi error set hui ho to vo clear ho jaye

Like this:



Yaha pr declaration wali baat kyu: Modern c ka compile kahi bhi declare krne deta he but humko phli line pr hi declare krna he.

NExT: Think for csv write and all

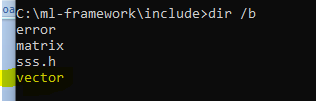
Write me issue : heading dena he to kya

h

Ab Vector banana wali baat:

Line page no. 26 se start he part 3 wali file me

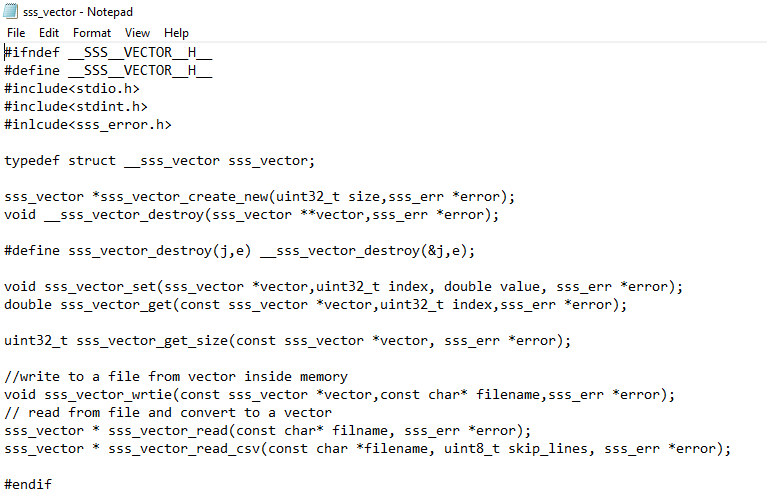
1. Include me vector naam ka folder



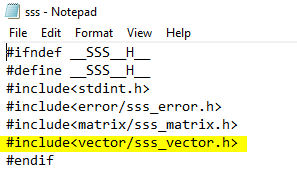
Usme sss\_vector.h header file



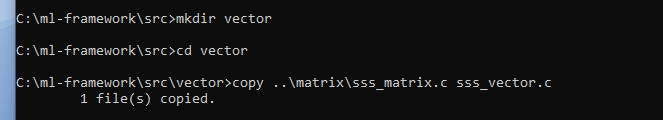
**code of vector ki heder file** :



Sss.h main header file me file me is sub header file ki entry



1. Ab vector ki source file bnate he



Humnesrc lkh li ss lena he then 36 se start krna he

Place ss of vector functions. And testcases

//pending come come

For completing the vector file code

As of now we are moving to the ML

**ML Start:**

Sample data k liye data set download krege.

Read and process and result

Dataset read kr k matrix me aa jayega.

Iske baad vector read and write wala code

Place ss for code and testcases.

Ye ho gaya he ss baki he

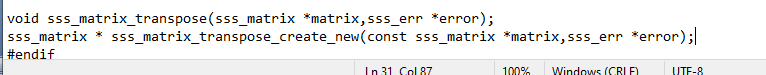
Ab transpose use kiya he vector read me : kyuki user file me kese bhi likh skta he data like

1 row and multiple columns ya 1 columns multiple rows

But humko to finally 1 column multiple row wala bnana he.

To uske liye transpose ka use

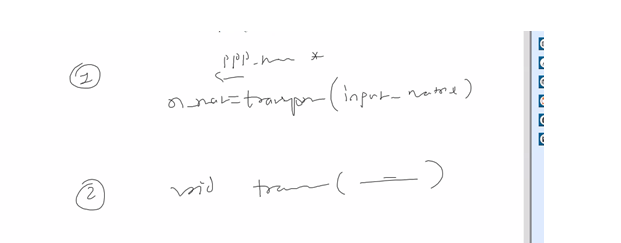
Transpose k liye matrix.h me 2 function



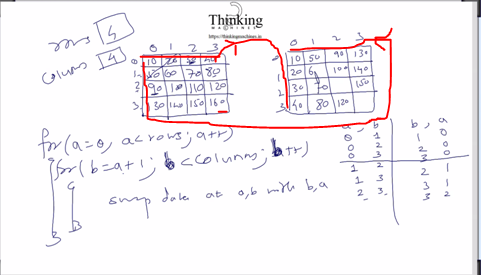
Page 50 se start krna he part 3 k ss ss folder me he and then part 3 khtm

Iske baad humare paas transpose ka code and

Vector read write ka code jo hum likh chuke means transpose ka code likha and part 3 end then move to part 4



Squere matrix k transpose ka logic:



Yaha pr (a,b) ko (b,a) krna he to transpose ho jayega.note: for squere matrix

Ye pura algo step by step run krna he.

Now:

Pick data at (a,b) .

Value=matrix->data[**a**\*matrix->columns+**b**]

matrix->data[**a**\*matrix->columns+**b**]= matrix->data[**b**\*matrix->columns+**a**]

(b,a) ka (a,b) ko assign kr diya

And then

matrix->data[**b**\*matrix->columns+**a**] = value;

now rectangular wale k liye:

iske rows uske column and iske column uske row

new\_row= matrix->columns

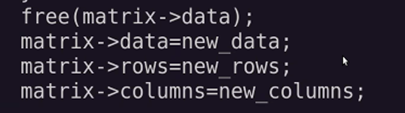
new\_col=matrix->rows;

new space = (double \*)malloc(sizeof(double)\*new\_row\*new\_col)

ab existing k (a,b) ko new k (b,a) me copy kr do



Then reassign the new config and clear the existing



2 type ka function likhna hoga,

Ek to in memory transpose, dusra create a copy of transpose. And return it.

Place ss for transpose code and testcases

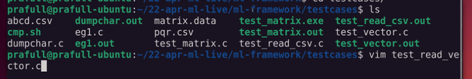
Transpose wala bhi likh diya matrix transpose wala.

Matrix print and vector print ka code.

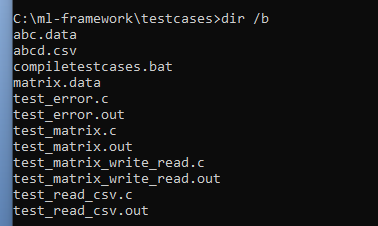
Ab matrix.h me ek function bdate he print ka and .c me implementation

Same for vector

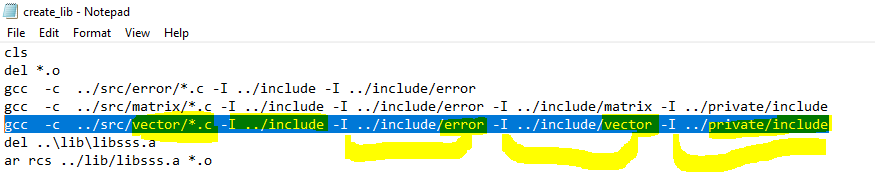
List of testcases by sir



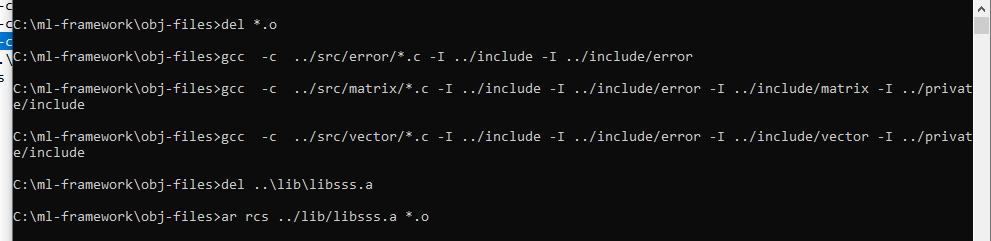
By us



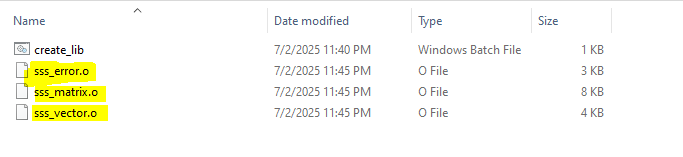
Abhi test krne se phle lib bnana , but vector k source ko compile krne k liye line add krni pdegi.



Ab is script ko run kiya and lib ready



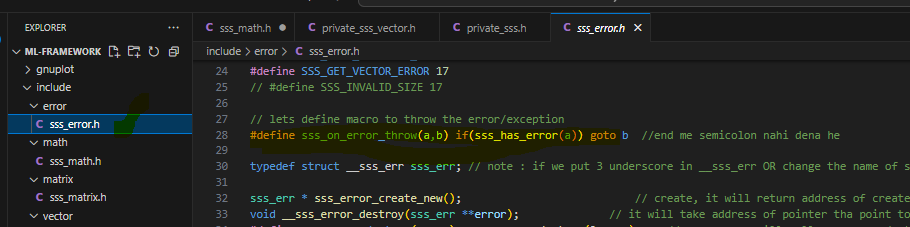
Now we have error.c, matrix.c, vector.c ki object files he in lib



Is lib ka refrence dekr testcase run krege.

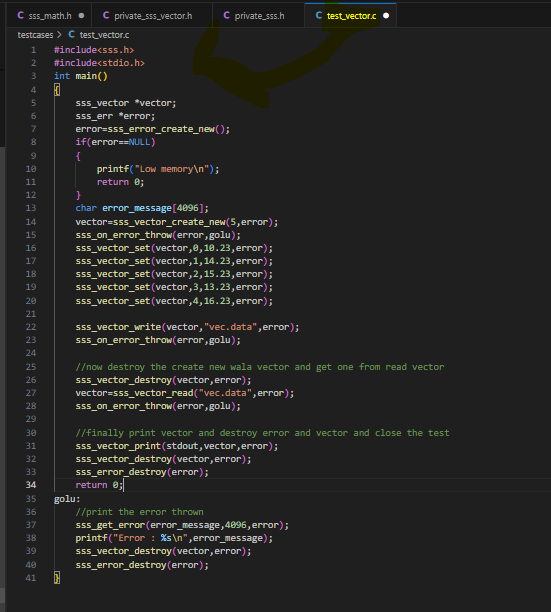
Lets write testcase in testcases folder

Iske phle ek goto k liye macro define krte he to throw the error:



Then compile it and create lib, don’t think , .h me change k badd lib create krne ki need he , vo to only .c k liye he.

**Testcase of vector.c:**

****

Ab bs ss lgane he and matrix me kuch row colum and dimensoion wale funcioin k ss phir se lgane he

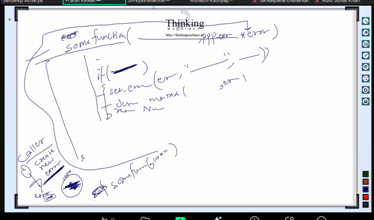
Now part 3 ends we can move to part 4

Lets move to part 4:

**Part 4 Starts**

5/24 may 24 fare July 4, 2025

Seterror me issue:



User call function A() and something went wrong he set error and just on the next line he call any other method B() and pass the same error object refrence to B()

But B() ran successfully and user checked the error after B() then he will het haserror() true.

To ye galat he :

**Solution** : in every function jisme error ka refrence as argument aa raha he usme clear the error after variable declaration.

abhi yaha se machine learning chalue he

**MACHINE LEARNING STARTS**

**Guideline:**

Aage kya: part 4 me:

1. Data load krna dataset me [function read csv], then split into input matrix and output vector
2. Ek matrix and ek vector bnana [create new()]
3. Matrix me first colmn me 1 fill krna then dataset k x-1 column ka data copy krna [function fill(), and function copy()]
4. Vector me dataset ka last column copy krna [function getvector from matrix]
5. Ek model vector bnana jiski size input matrix k column k barabar ho [create model]
6. Model me randon m=n
7. o. fill krna [function fill random()]
8. Ab input matrix amd model ka mulriply kr k ek vector taiyar krna [function multiply matrix with vector()]
9. Ab ye vector and vo y original vector ka difference nikalna [function diff()]
10. And unka sum kr k mse nikalna then model ko update krna bsed on ki y fit hua k nahi

Kiya kya he:

Y=mx+c , m and c ki kuch value le k ek line draw kri

Means ek dataset ready kiya

Ab ye data set training data set bna ,

Ab humne randorm m and c liye and is data ka x diya and bola ki aapke random m and c + is dataset k x ko formula me rkh k jo y aaye vo original y k equal hona chaiye

Agar jis m and c k liye vesa ho gaya vo m and c ka jo model hoga vo trained model mana jayega.

To ye step me job hi function aaye he vo hi likhe he pure part 4 me.

Iske baad gnu plot ko install krna and cmd se data lekr graph draw krna.

Ab next

Read csv se data load then

Ab split krna he.

N input 1 output column

Ek matrix ek vector bnega.

Function jo add hoge:

Matrix\_copy()

Matrix\_fill()

Matrix\_get\_vector() // matrix se vector nikalne k liye

Fill copy and getvector teeno matrix me rahega

Ab ek or bnega multiply\_matrix\_with\_vector()

Ek multiply\_matrix()

Ab isko kaha likhe vector me ya matrix me.

To iske liye ek alage folder bnega utility folder

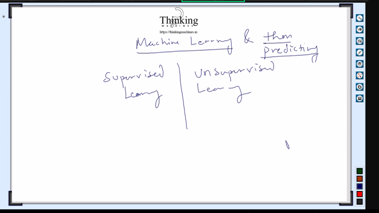
Ek or uitility

Mean\_squered\_error()

Then matrix\_write\_csv()

**Machine will learn and then start predicting**

There are two types of learing supervisod and unsupervised



Abhi unsupervised se koi mtlb nahi , humara bahut sa samay lgne wala he supervised me.

Ye kaam india me bhi hota he .

Robotic: separate the product between defect and non defect

Bahut sari MNC’s me model train krne kaa kaam hota hi nahi he.

Vo to openai ko request send kr k response lete he.

openAI is also a model , do a deep study.

Note: it’s unimaginable ki 11th 12th ka sara math use ho raha he AI me kahi na kahi.

Now what technique we need to use to train model?

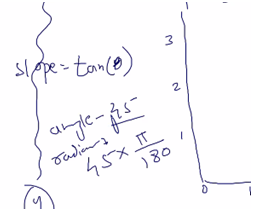
Ans: depend on ki data kesa he.

Let’s take and example of linear regression.

Step 1: install gnu plot for windows, it is very powerful tool

Now y=mx+c where m is slope that is **tan theta.**

**And c is the intercept: a point on y axis where line cuts, means is always 0 at that point (0,y) is the point.**

****

Programming me theta ko radian me convert krna pdega.

Download gnuplot exe and install it , it will goes to this folder



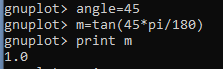
Now run the program: "C:\Program Files\gnuplot\bin\gnuplot.exe"



We can invoke gnuplot program using java as well.

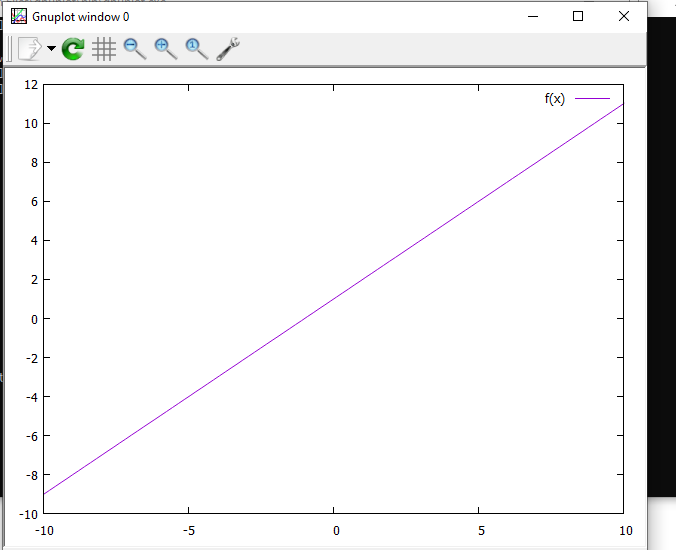
Now using command from gnuplot , we will tell gnuplot program to draw a line step by step:





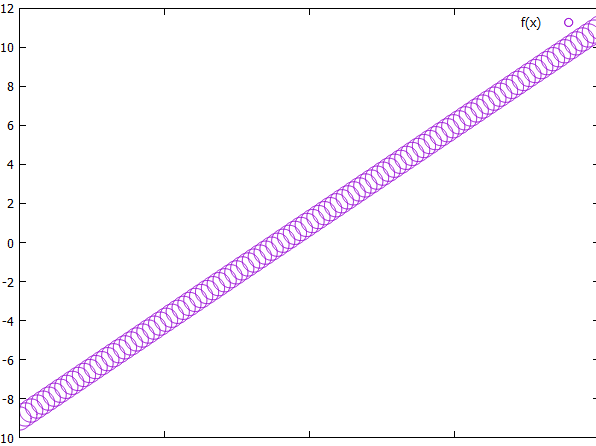


And esa window open hua





Is se esa bna

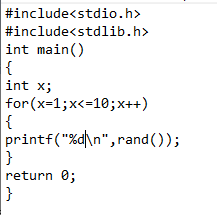


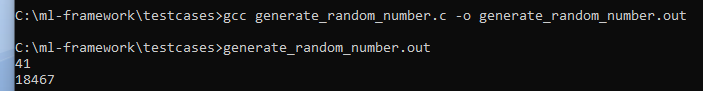
Step2:

Ab programmatically data set generate krege:

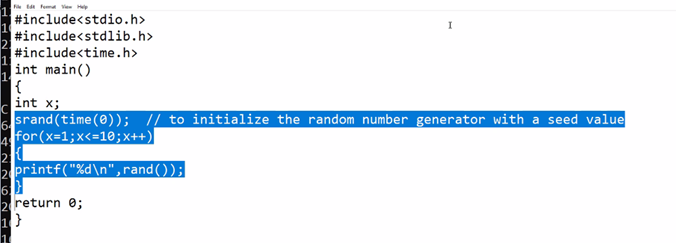
Rand() k liye stdlib







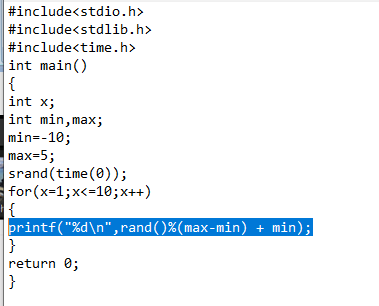
Ab isme sequence generate ho skta he to ab randomness daal dete he or bhi



Time.h se time() ka use,: This ensure a separate sequence every time.

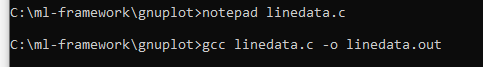
Note : time() ko 0 ya NULL bhi pass kr skte he.

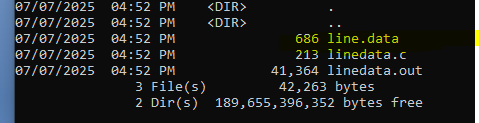
Ab kisi range k bich ka number generate krte he.

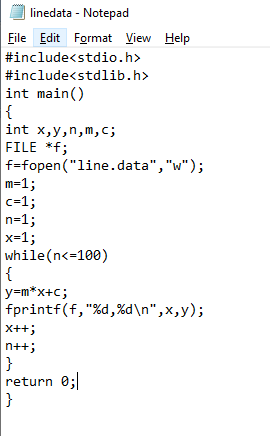


Is formula se jo bhi number generate hoga vo range me hoga. Min max ki.

Ab ek c ka program likhte he jo ek file bna k x,y k pair daal dega







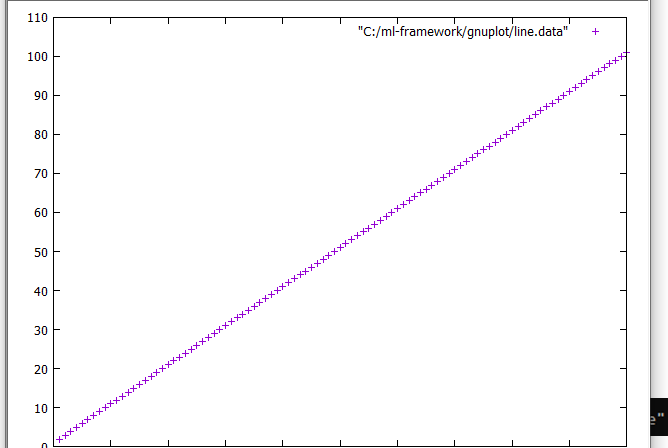
Is program se data file me likha jayega.

Ab is file ka use gnuplot me kr k line draw krte he.

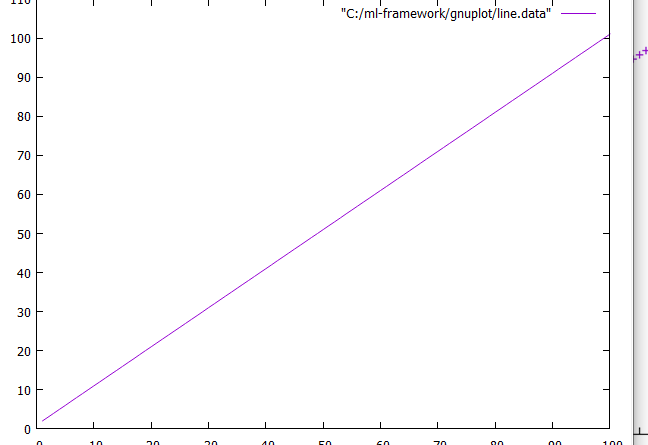
"C:\Program Files\gnuplot\bin\gnuplot.exe"



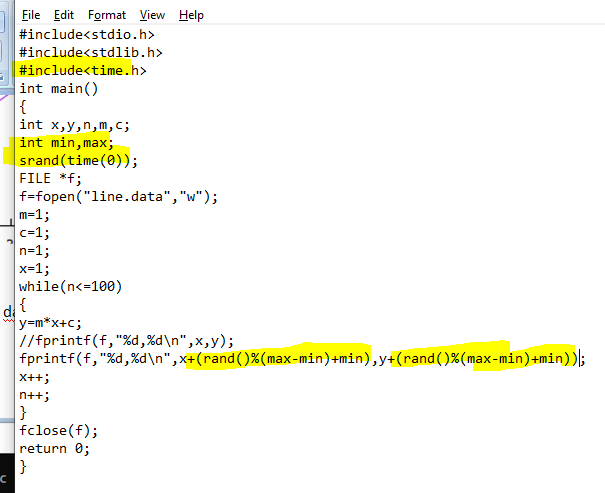


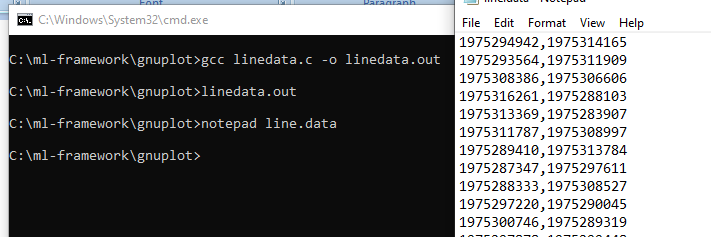




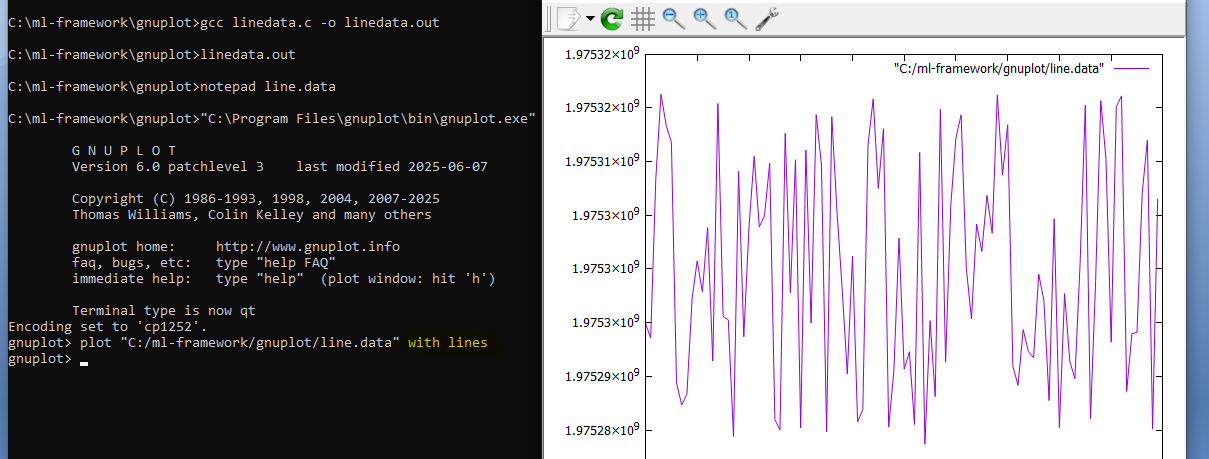


Ab line.data me random x and y dalte he

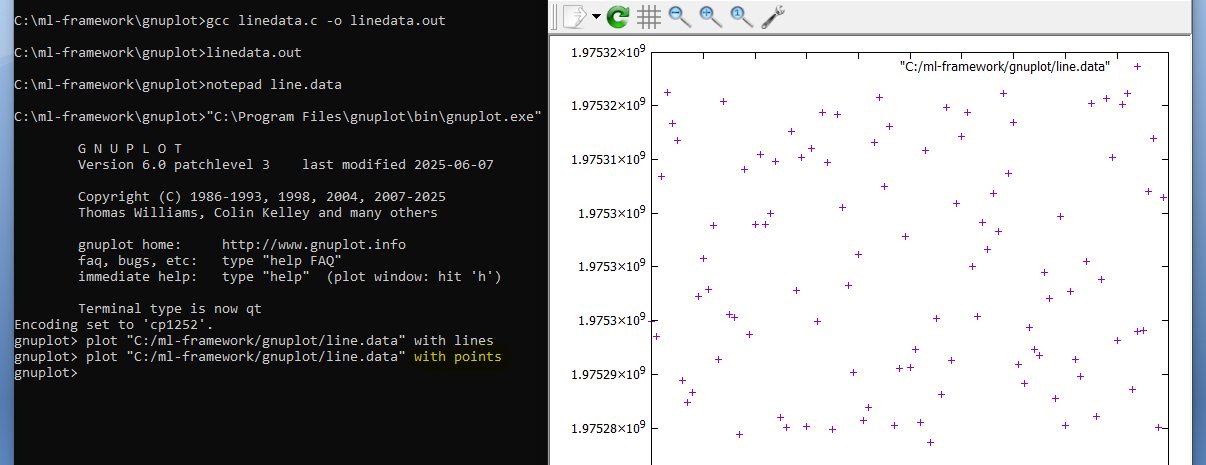




Ab is data se dra krte he line:

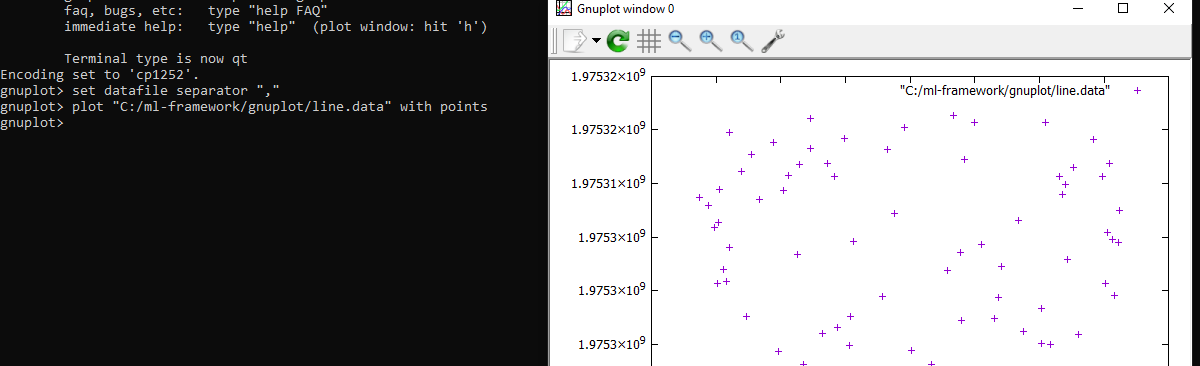


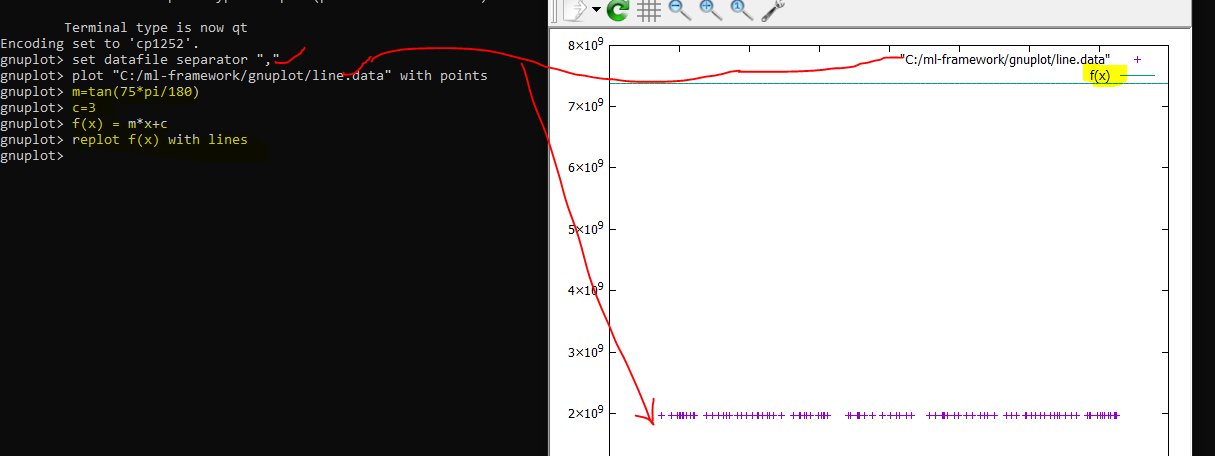
With points:



Ab is bikhre hue data ko align krna he.

Lets take m and c with all points and then redraw. It





Here gnuplot ko btaya ki x,y are separate by , in file the plot the data

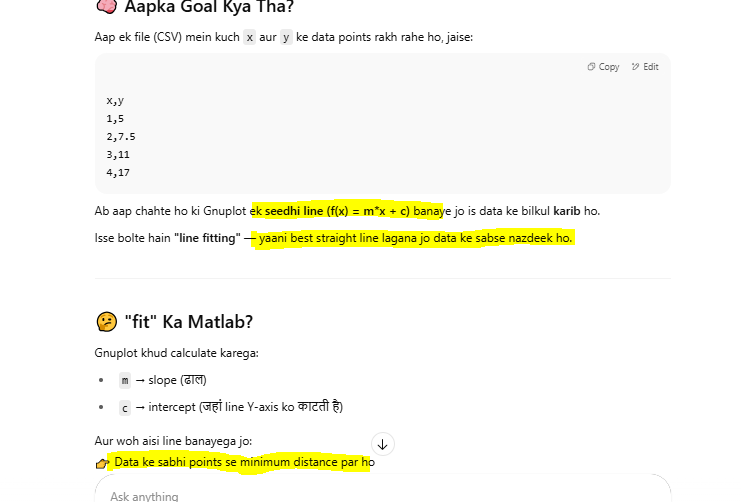
Then set m ,c and create a fx then draw this f(x) : here m and c are random

Now we can see output of line.data and output of f(X) both , In this graph.

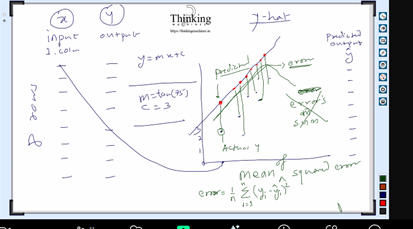
Now what is our goal?

Ans: m and c ki esi value find krna , jis se esi line bne jo ki given data k sbse pass ho

Means saare data points se minimum distance pr ho.



Jo humara line.data and f(x) ka jo graph bna , uske according dekha jaye to Tend towards , below



Ab humko kya krna he.

File se data read kr k matrix me dalna he , x and y k 2 column bnege where y Is the actual f(x)

Ab hum random m,c and matrix se x lekr y’ or f’(x) ya bol skte he naya y find krege.

In saare x k liye aaye y’ ko ek vector me dalege. Jiski size no. of rows in matrix ya no. of x [data point].

Now after one round Ab humare pass 2 vector he y and y’.

Ab error vector nikalege E=(y-y’)

Since negative value ho skti he to isko squere krege.

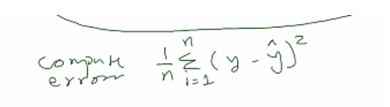
E^2 = (y- y’) ^2

Ab is E^2 wale vector ka sum compute krege.

To formula summation of (y- y’) ^2 from 1 to n

And average nikalne k liye divided by n.

To esa bnega final formula



Yha humara goal rhega error ko 0 krna.

To phir se loop with new m and c with same set of input data x. []

Ese total N number of loops . [update m and c on each iteration]

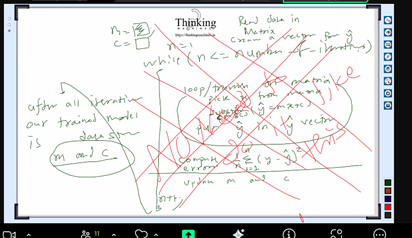
Jb saare trial ho jaye tb jo final trial me jo m and c rhega vo humara trained model rhega.

Vo value gnuplot ko dege to vo esi line draw krege jo line.data k saare point k paas ho.

Ya line.data k har point ki distance is new model wali line se minimum ho.

This is called trial and error method.

Tarika sahi he perfect he but ye tarika nahi use krna.



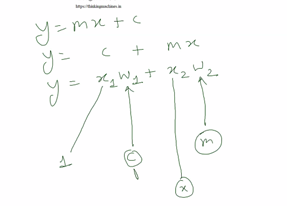
Ye gaya kachre k dibbe me.

Kyu galat he? Kyu ki koi bhi ml ki team isko use nahi krti.

Ab naya tarika:

Y=m\*x+c ko ese bhi likh skte he.

Y= c \*x1 + m\*x2 where x1 =1and x2= x;



Phle ek input tha x , ab humne ek or bda liya 1.

Ab hum 2 input ka ek marix bnayege and har input ka koi coefficient hoga like m and c ye kl ko multiple inputs hue like x1,x2,x3…. To unke liye m1,m2,m3…. And ek c and dummy x as 1.

Filhal 2 input he to 2 size ka model vector.

Means jitne inputs ya jitney input matrix k columns , us size ka model vector.

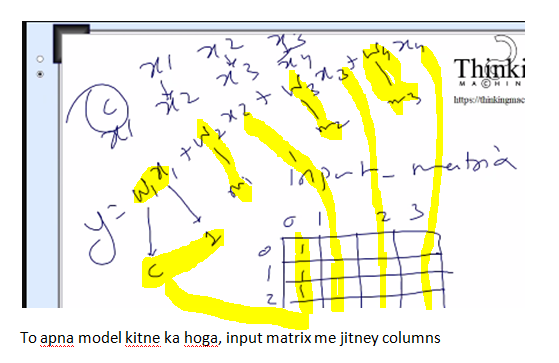
Ab is matrix ki har row me ek data rkh do and har row ko model vector [column vector] se multiply kro to esa aayega:

X1m1+ x2m2 +x3m3 + x4c….. and ek scalar value hath lgegi.

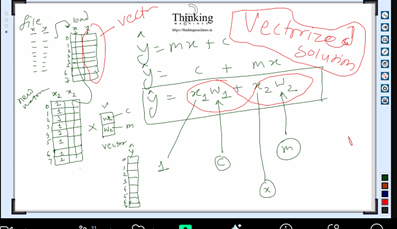
Is scalar value ko y’ wale vector me dalte jayege.

Jitne input utne m and ek c with x as 1.

Jitne input utne m isliye ki har point ka ek slope hota he.



And esa kr k after multiplying each row of input matrix with model vector we will finally get the y’ vector.

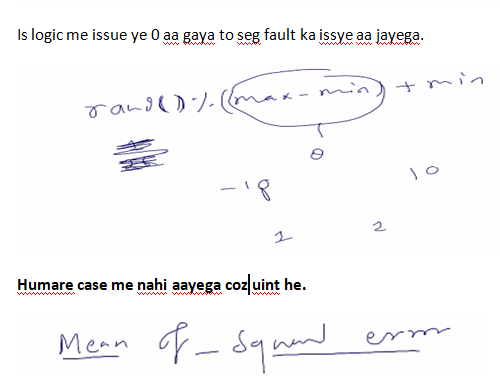


**And this is called vectorized solution** : if solutinon is vectorized he to parallelism k liye library available he

Gnu plot bnane wali company hi vo library likhti he.

Ab next: jo random() number ko range me generate krne ka logic tha usme seg fault aa skta

He.



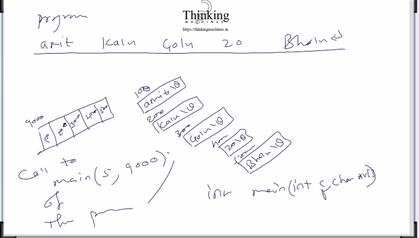
MSE : mean of sequered error [here this errors are list of errors from each point predicted and actual]

Next: main function me ek esa code likhna he jis se trained model mil jaye

Filename will be linear\_regression.c

It will be in testcases folder

Before this we need to learn how to accept cmd args in c



In above ss, user passed 5 argument separated by space , each argument will be considered as string and system internally will create each string and terminated by \0 and put the base address of each string into an array and then pass base address of this array that is 9000 in our example .

This array contains 5 pointers of char \* type.

So we can capture this in array to pointer to char , that is array of char \*

Int main(int count, char \*v[]) , Int main(5,9000) (total no. of args, pointer to array of char \*)

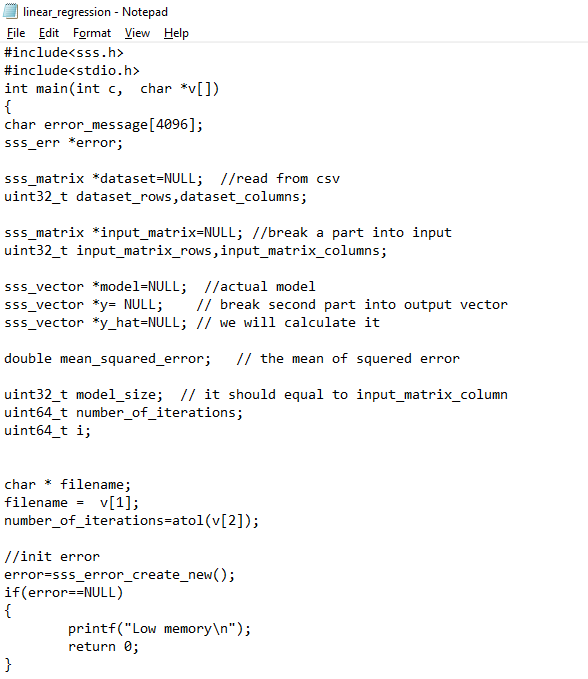
V[0] isprogram name , v[1] se argument chalu hoge , kitne hoge vo 5-1=4;

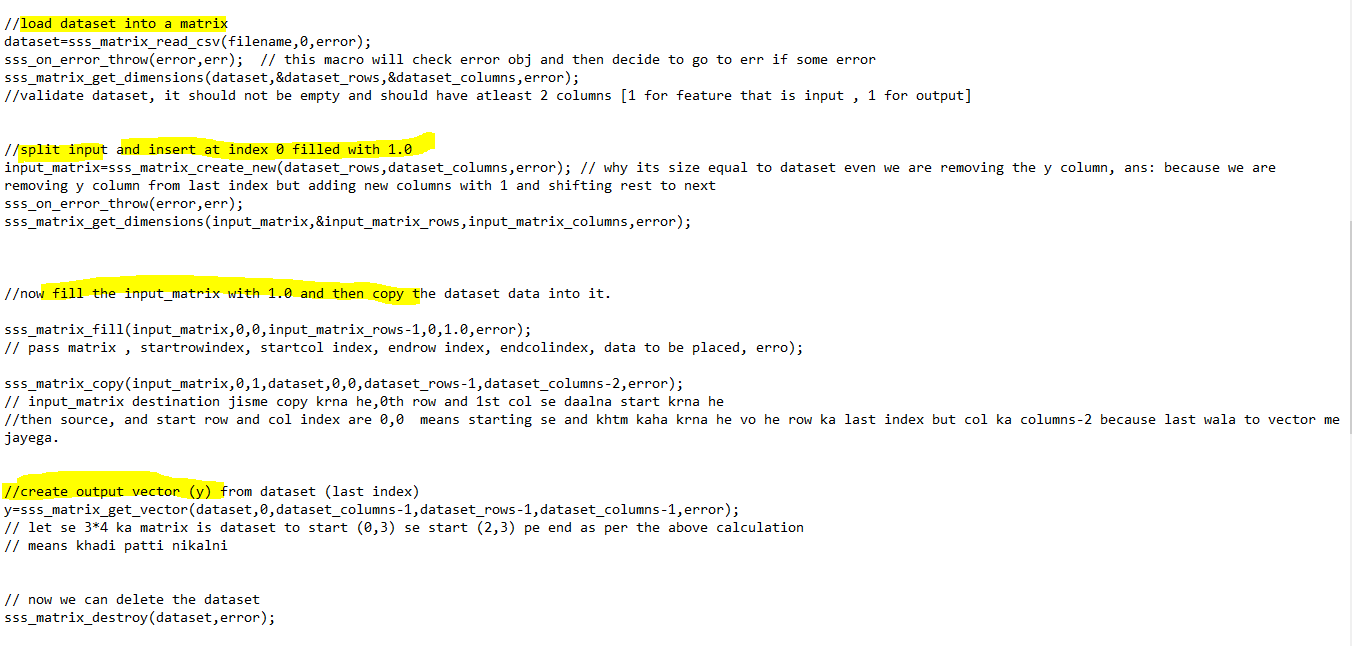
Now lets create a file linear\_regression.c

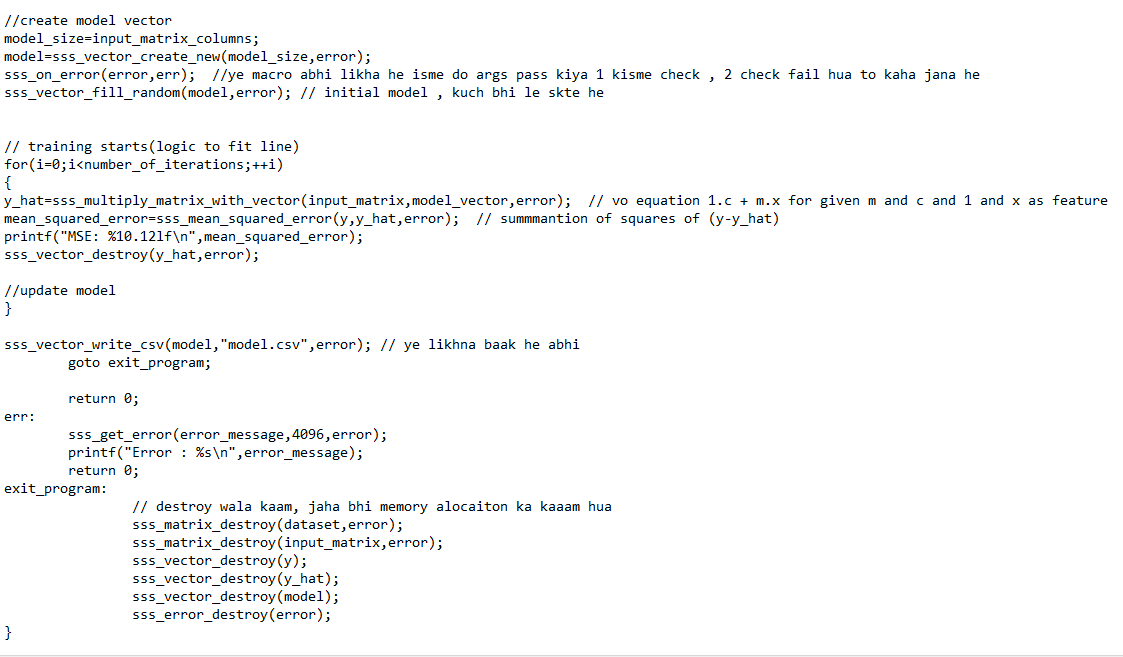


At page no. 20 we are in part 4.

Ss se bnaya code:

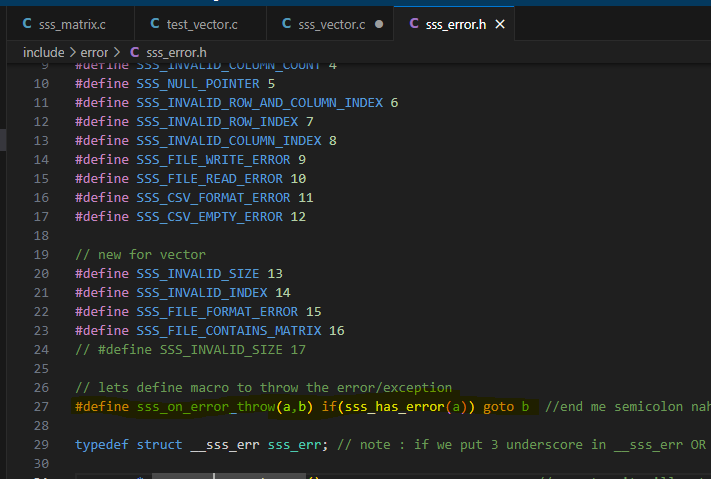






Note: add error check after creation of yhat

Define error throw macro in error.h file

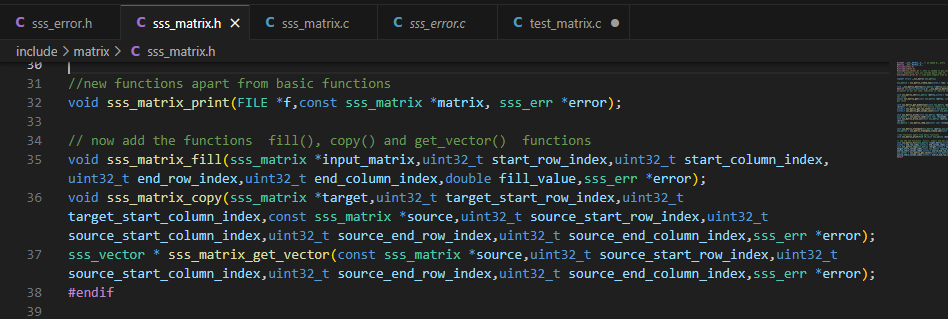


Now define each function used linear\_regression.c

Now matrix k functions copy(), fill(), get\_vector() bnate he



Added this three function in .h file



Let say start point is (x,y) end point is (p,q) for a matrix with r\*c then

Then

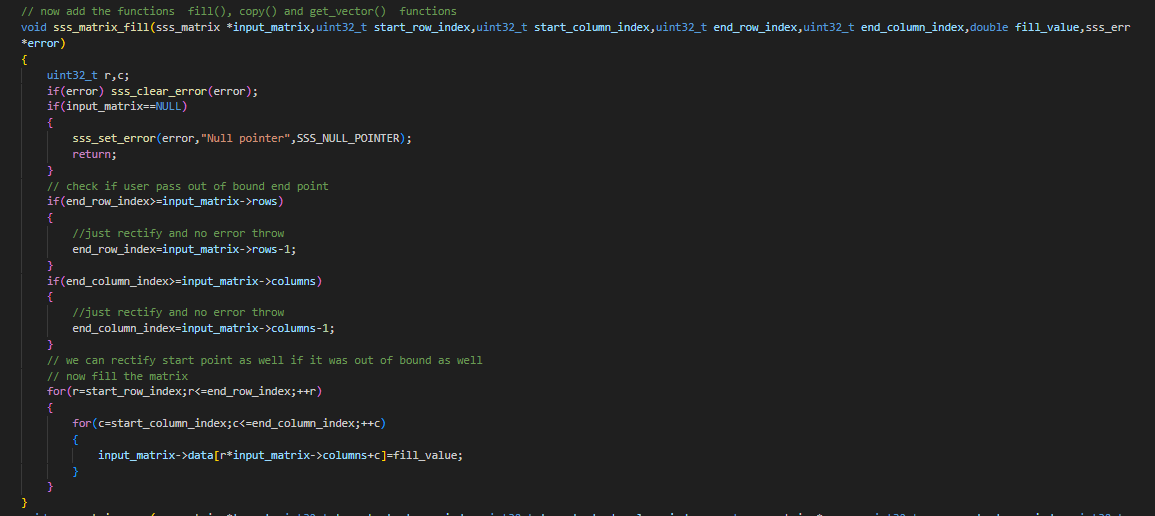
x : start\_row\_index

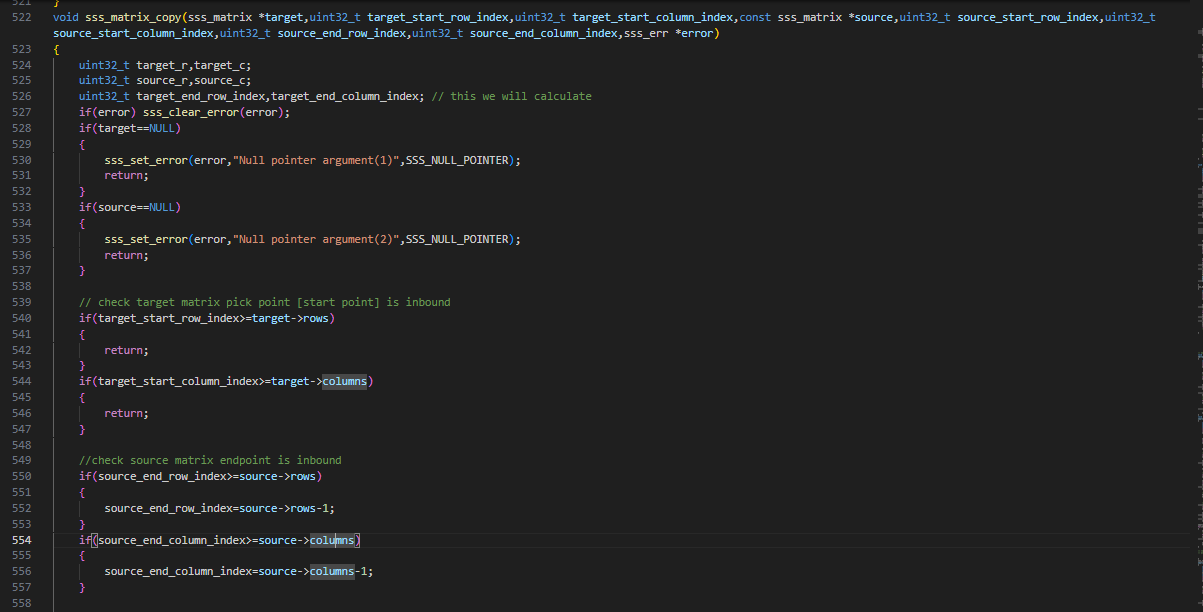
y: start\_column\_index

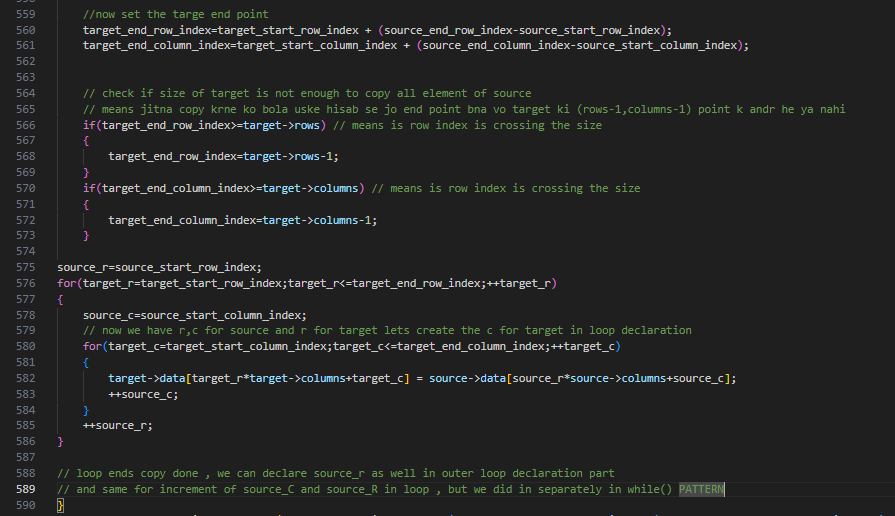
p : end\_row\_index

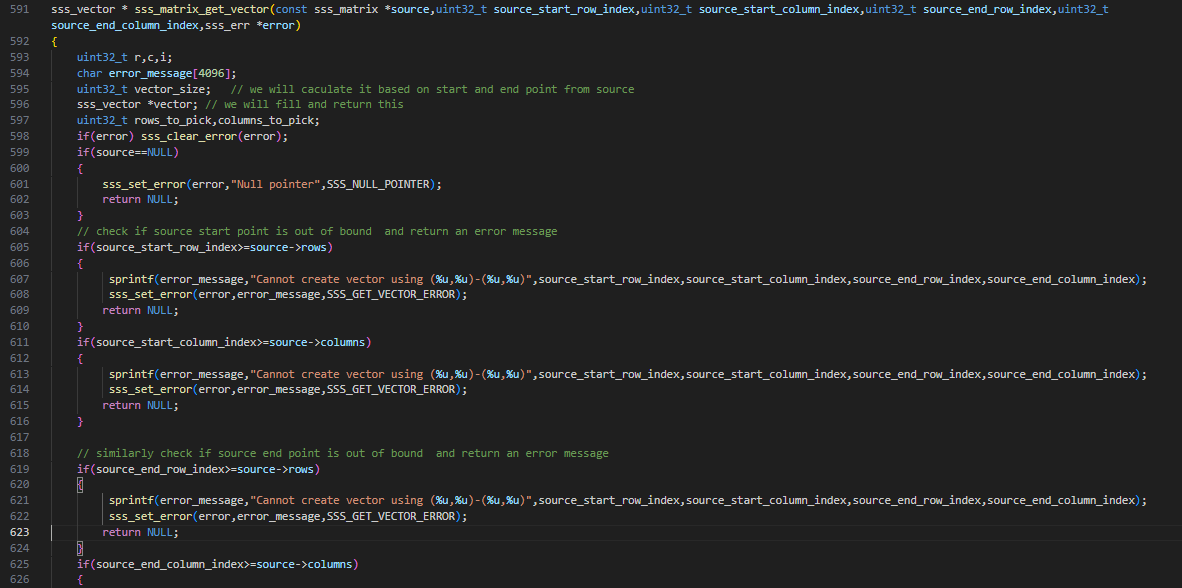
q : end\_column\_index

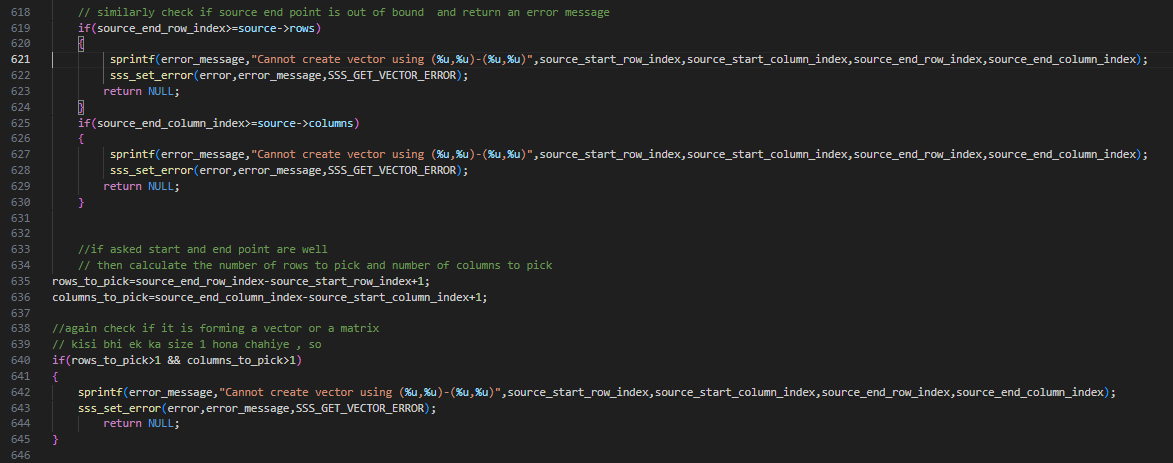
lets define the function in source file:

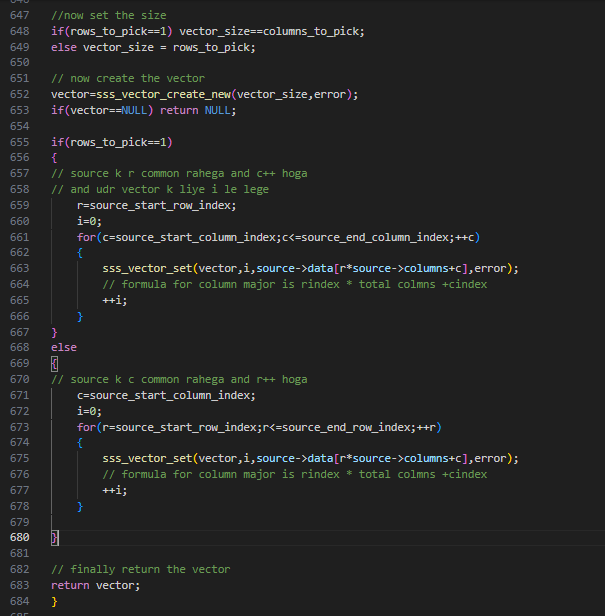






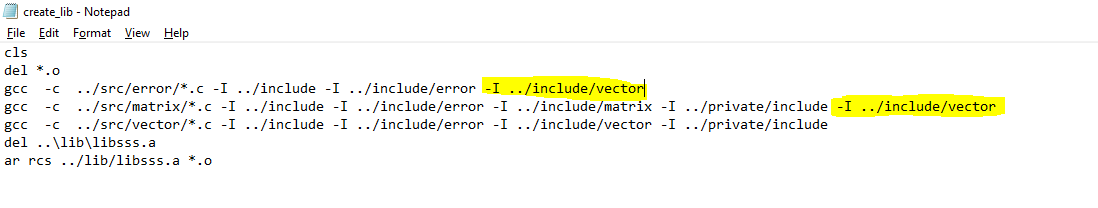






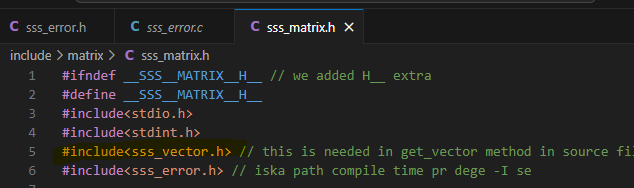
now recreate obj file

Creat lib:



Isme error ki source file ko compile k time vector ko kyu include kiya?

Because error me sss.h he and usme matrix he and usme vector ko include kiya he , for



Here we included vector for below line

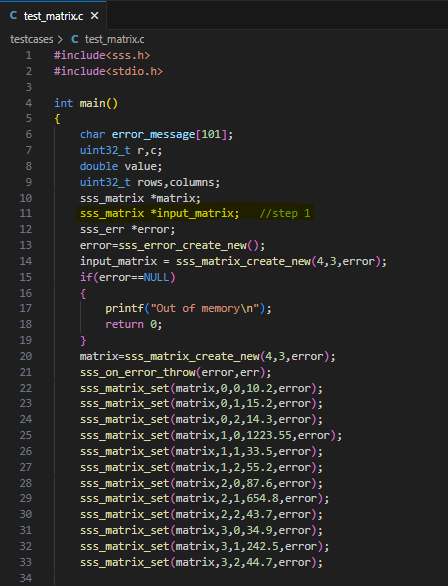




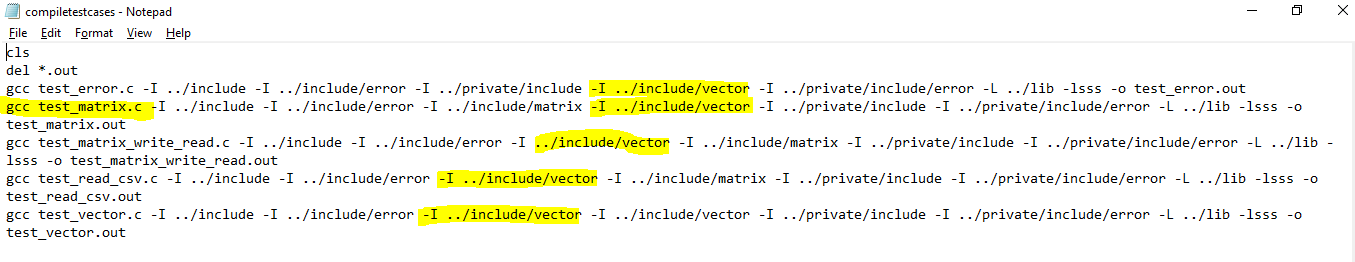
Iska testcase bnate he:

Matrix fill() copy() and get\_vector() ka.











Note: every time when there is change in src then create lib then compile test case and then run it to see the latest change from lib.

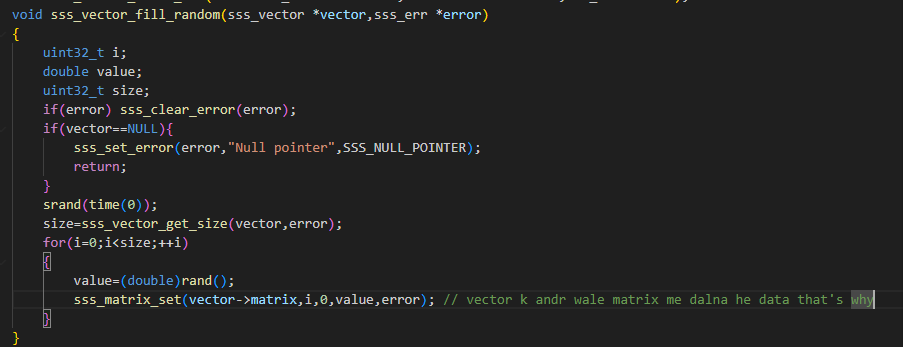
Next:

vector me ek method bdegi fill random()

**Vector.h me:**

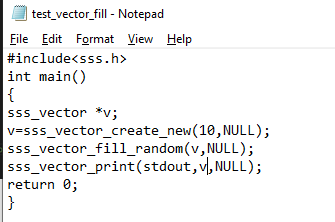


**Vector.c me:**



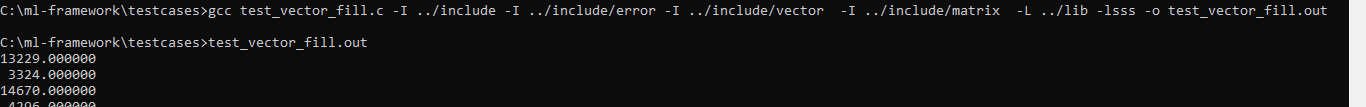
**Vector.c vector\_fill() testcase:**





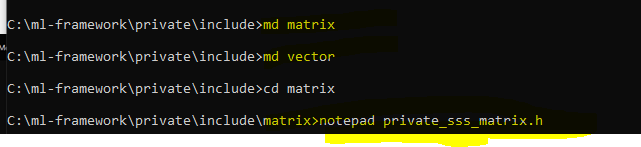
gcc test\_vector\_fill.c -I ../include -I ../include/error -I ../include/vector -I ../include/matrix -L ../lib -lsss -o test\_vector\_fill.out

the above line is to compile

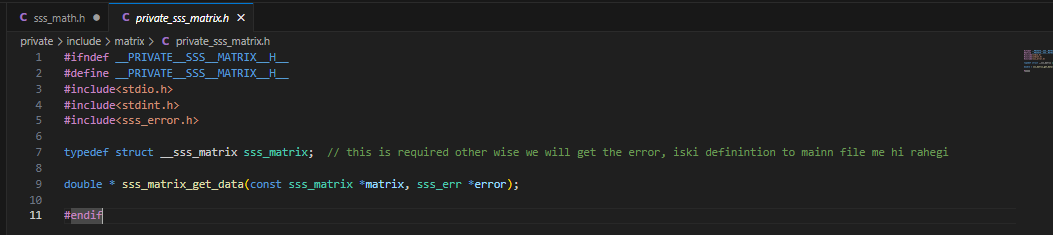


Now: since we need data from matrix , so we are going to create an private matrix file for internal use.

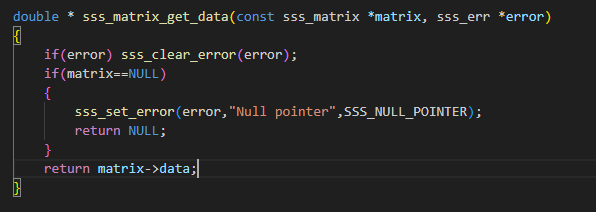
Same for vector as well



**Private matrix.h me:**

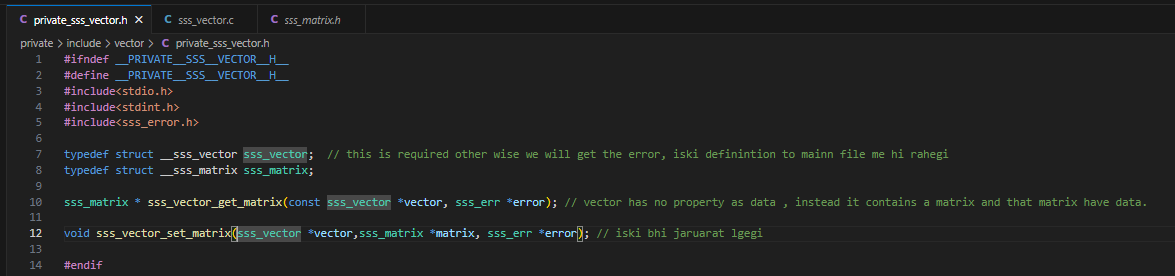


**Matrix.c me:** get data k liye

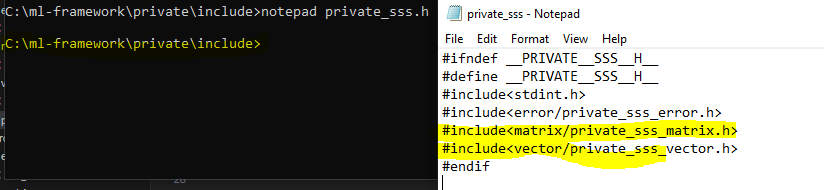
****



**Private vector.h me:**

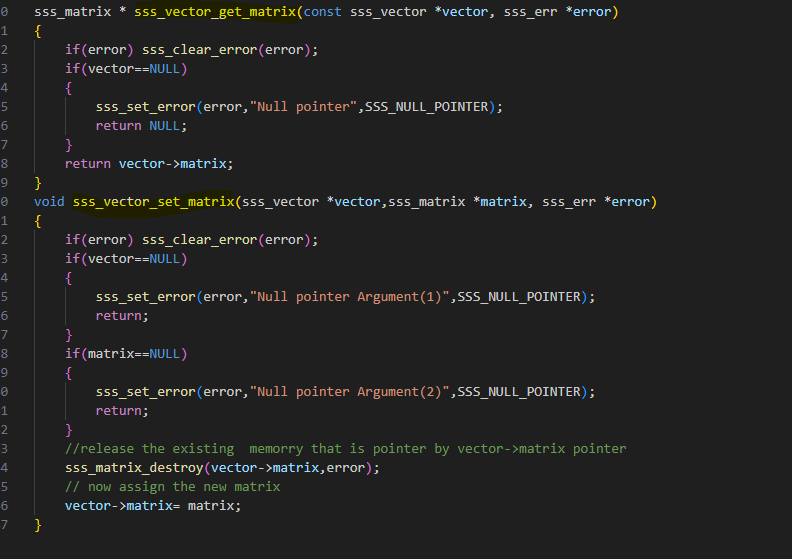


Now combine this private header file to main private header file.



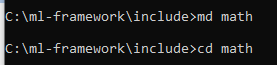
Ab inko source me define krte he:

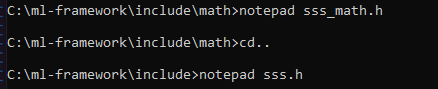
**Vector.c me:** get matrix k liye

****

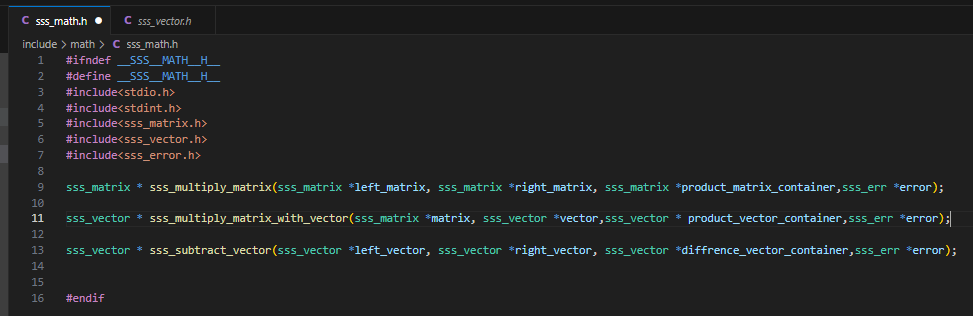
**Ek header file bnegi math naam ki , jo ki math operation k function rahege usme**

Now math operation k liye ek alag header file bnegi jo ki math contain kregi



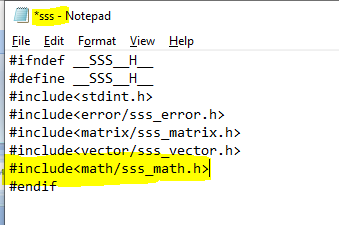


**Math.h me:**



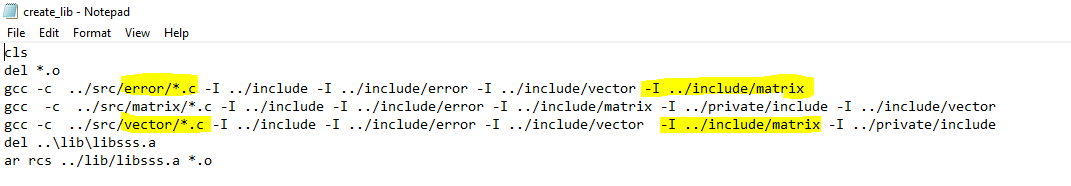
Container isliye liya kyuki ye functions loop me chlege and every time ye result object bna k return, then loop me kaam hone k baad release , to un sb se bchne k liye container liya[ek memory jisme bar bar rkhte jayege]

Ab isko main header file sss.h me include krne ka kaam



Now since we included **math.h** in **sss.h** and this sss.h is included in src of vector matrix

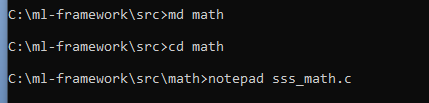
And at compile time it will open math.h and try to find the path for matrix.h from math.h but it will not get any such path so we need to specify this is createlib.batch file as below

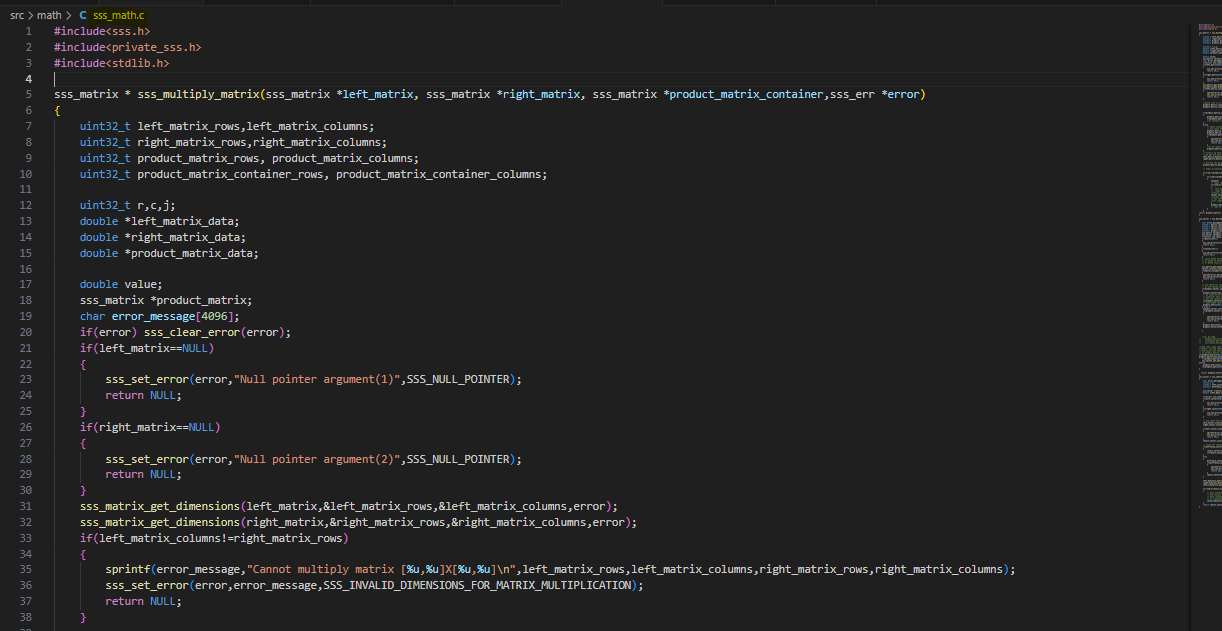


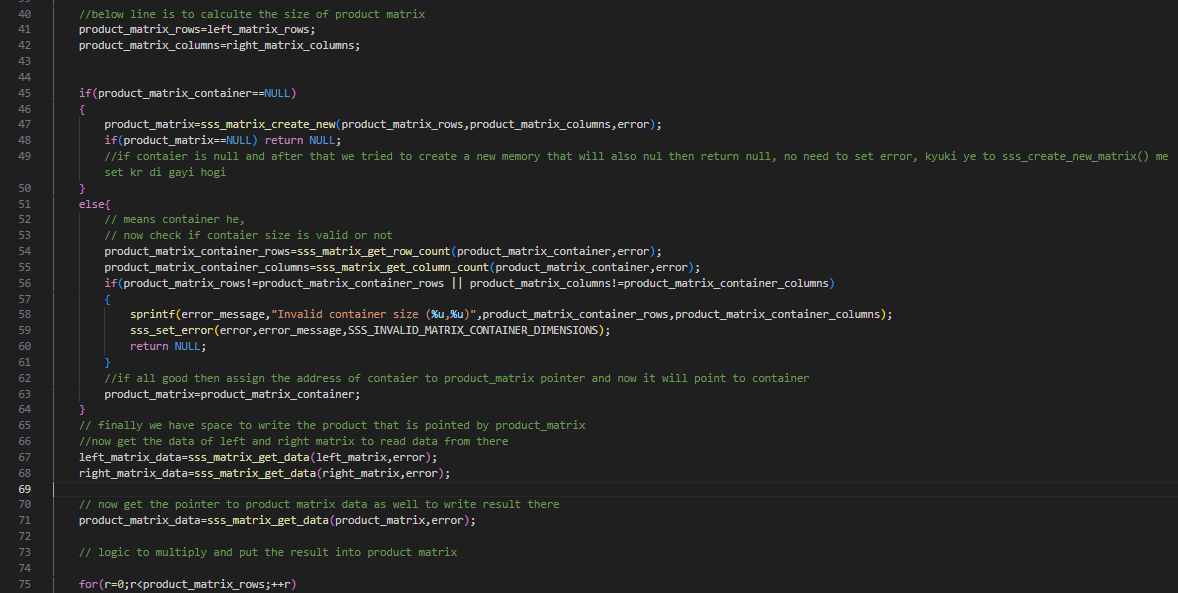
Matrix\_multiplication algo discussion:

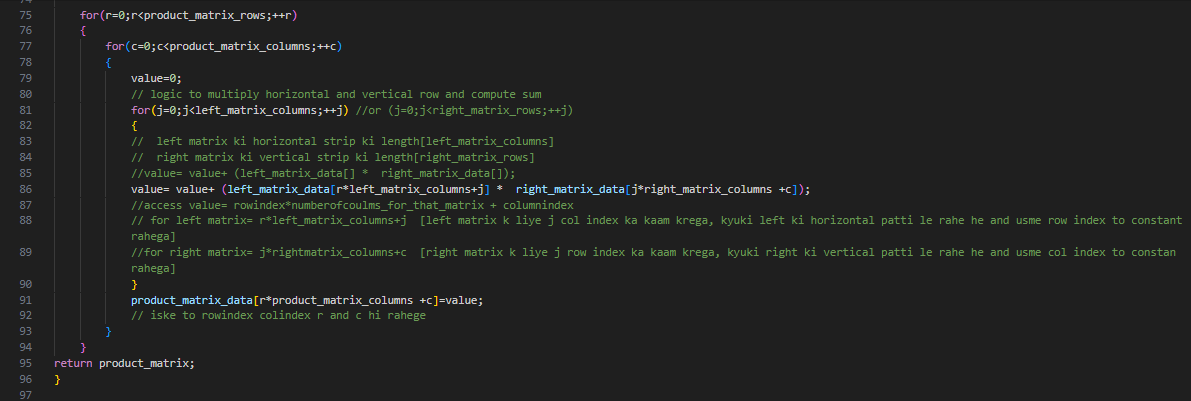


**Math.c me:**

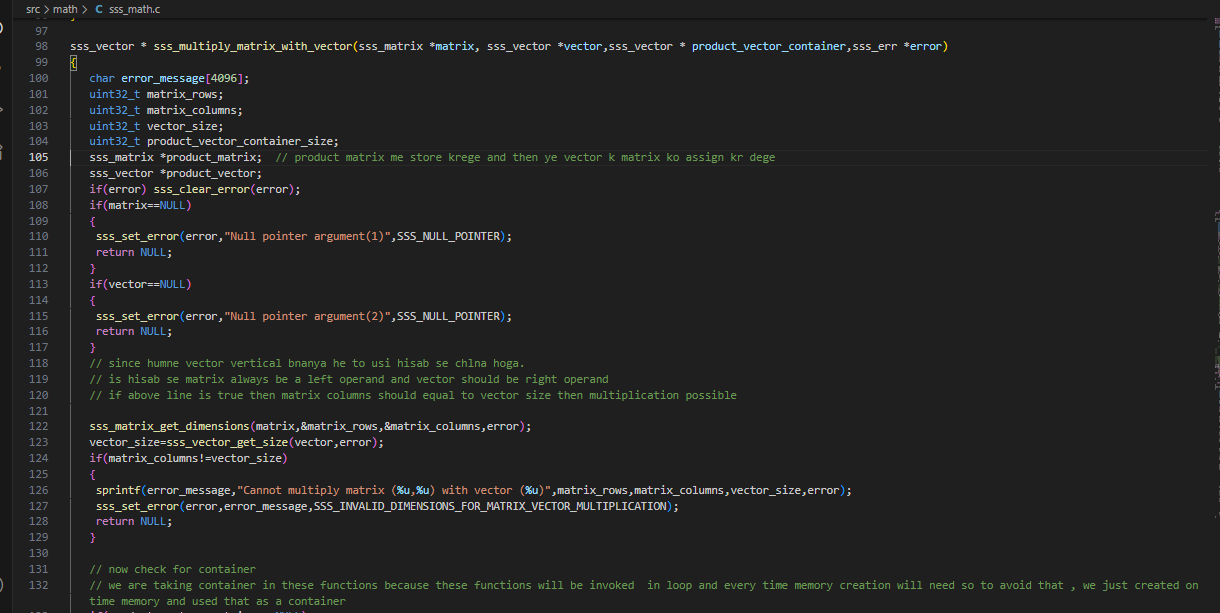
****

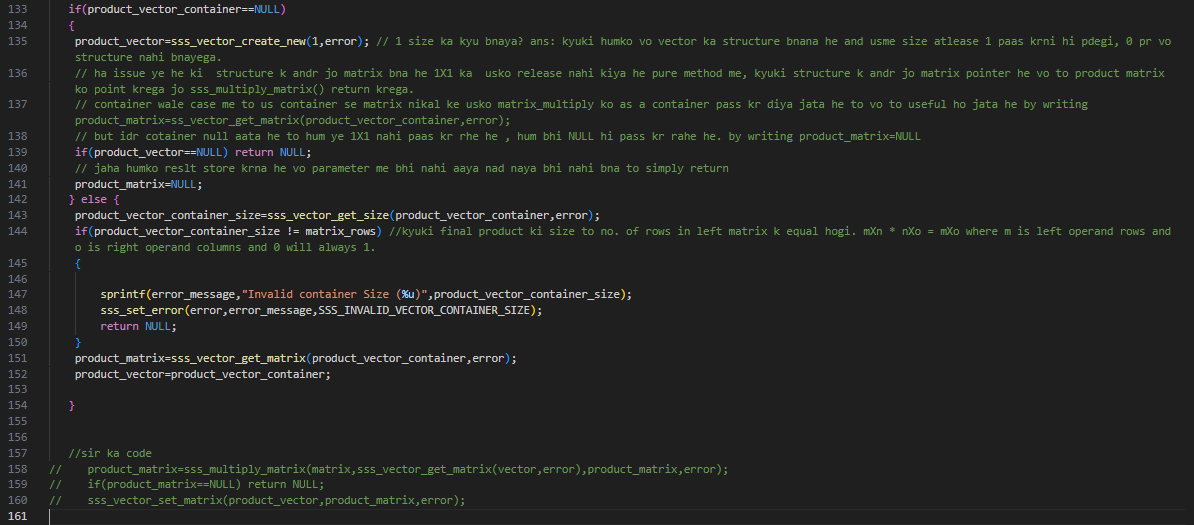
****

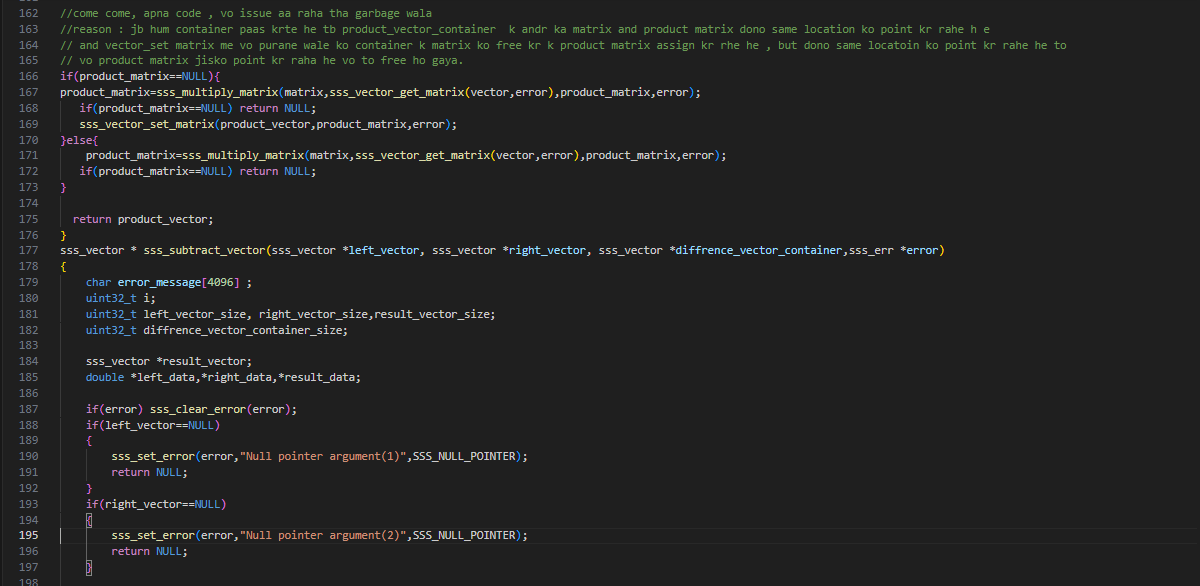
****

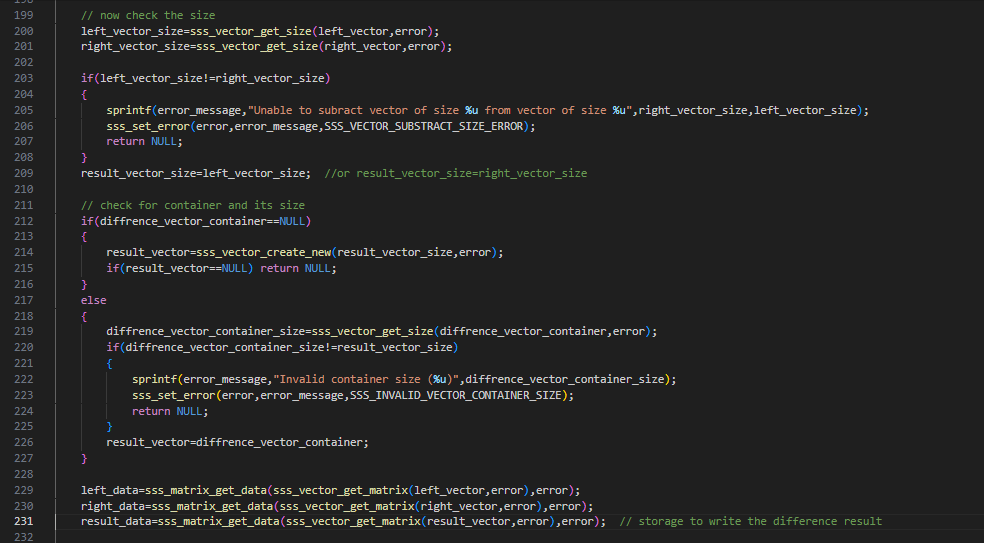
****

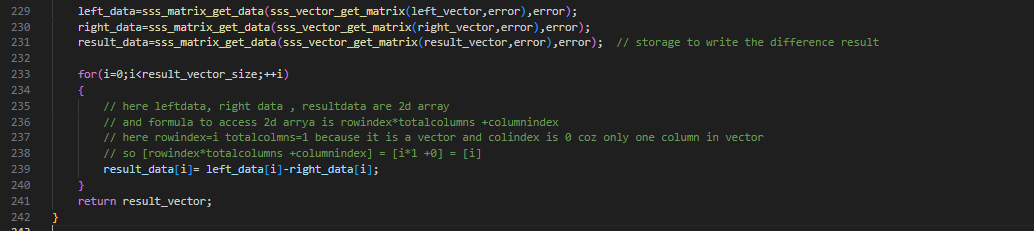
**Mudmultipy matrix done . now matrix with vector:**

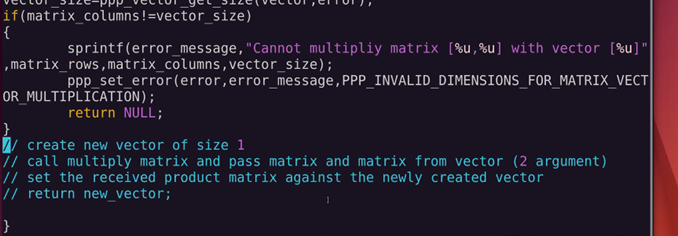
****

****

****

****

****

****

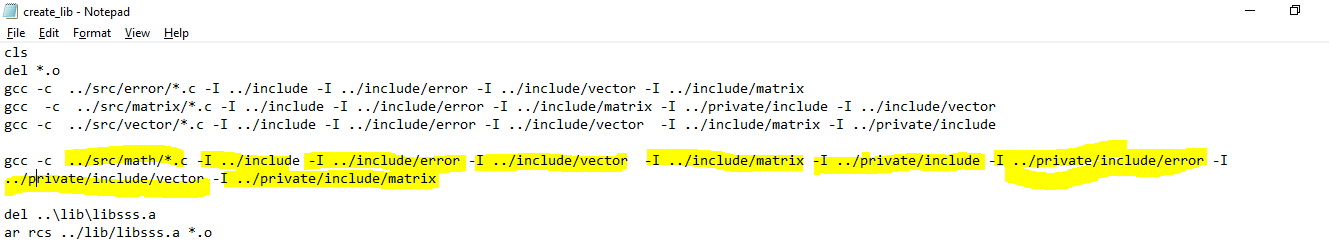
**Math.c compile k liye code:**

**.**since math.c me sss.h and private\_sss.h ko include kiya he to isko compile krne k liye iske andr

Sss.h and private\_Sss.h me jitni bhi header file include he un sb kaa path dena hoga

Like include folder ki error, vector, matrix

And private include folder ki error, vector , matrix files ka path.

****

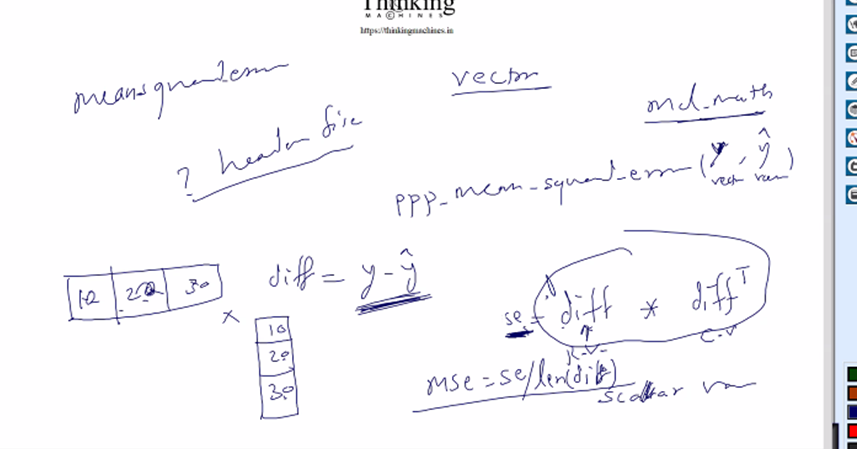
**Math.c k testcases: testcasefor this functions is writeen in file test\_math.c in testcases folder**

**Testcase k compile ka code:**

**gcc test\_math.c -I ../include -I ../include/error -I ../include/vector -I ../include/matrix -L ../lib -lsss -o test\_math.out**

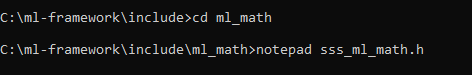
abhi hum part4 me page 90 me he : abhi fatafat code likhege and review baaad me krege aaaj 19 july 2025 ko ye code chlana he sync krna he

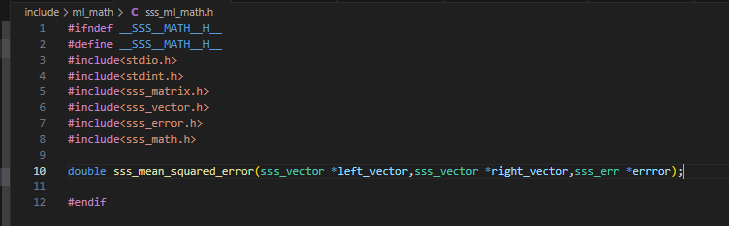
mean squered error nikalna he : (y-y)^2 , yaha kisi vector ka squere nikalne k liye uska uske transpose k sath multiply kr k sum kr do.



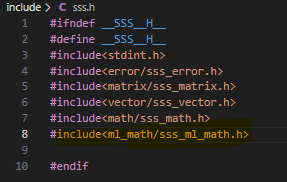
Ml k functions k liye ml\_math bnegi.







Isko main header file me include:



Ab iska source:



**Source ka ss lagana he.**

Isko compile krne ki line in createlib.bat me

gcc -c ../src/ml\_math/\*.c -I ../include -I ../include/error -I ../include/vector -I ../include/matrix -I ../private/include -I ../private/include/error -I ../private/include/vector -I ../private/include/matrix **-lm** -Wall

**note**: Math.h ka use kiya to c­pile k time –lm dena pdega.

Ab next: linear\_regression.c

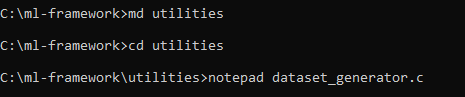
Uske phle **matrix\_write\_csv in matrix.c**

code likh diya he ss lena he.

We jumped to part 5, we are continuing here , later on we will separate it.

**Utilites bnate he**

Ab bnate he utilities:



Ye jo datasetgenerator he ye training and testing k liye random data generate krega.

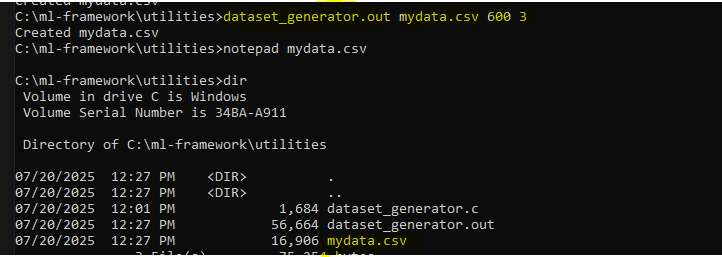
Iska code bn gaya he utilites me : isne matrix\_wrtie\_csv ka use kiya he

Isko compile krne ki line

**gcc dataset\_generator.c -I ../include -I ../include/vector -I ../include/matrix -I ../include/error -L ../lib\* -lsss -lm -o dataset\_generator.out -Wall**

iska ss lena he.

Isko run krte time cmd args lgege like this



Isko run kr k humne ek 600\*2 ka matrix bna k usko mydata.csv me likh diya he.

Ab isko split krne ka code: mydataset.csv ko



Iske baad split data into training and testing wali .c file jo percentage le k training and testing csv bnayegi. Ye bn gaya he

Compile:

gcc create\_training\_testing\_dataset.c -I ../include -I ../include/vector -I ../include/matrix -I ../include/error -L ../lib\* -lsss -o create\_training\_testing\_dataset.out -Wall

run:



Training me 419 and baaki testingme chli jayegi

Ab line k liye vo data generator wala data kaam nahi aayega uske liye to line ka formula use kr k bnana pdega

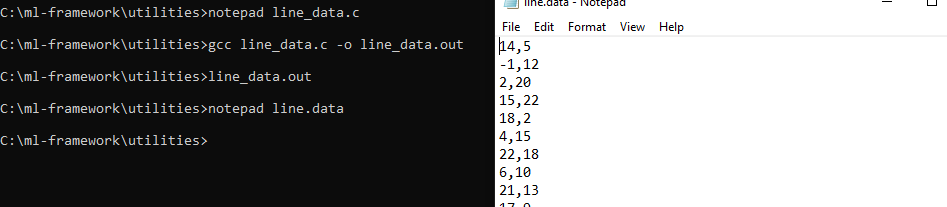
Vo data\_generator me to matrix k o random no. se bharne ka kaam kiya bs

Split krne wala program to kaam aa jayega.

Iske baad linedata.c bnayege .



Ye actual line k according data generate krega(x,y), ye ho gaya line ka data generate.



Iske baad diferential ka discussion : read thoroughly . :

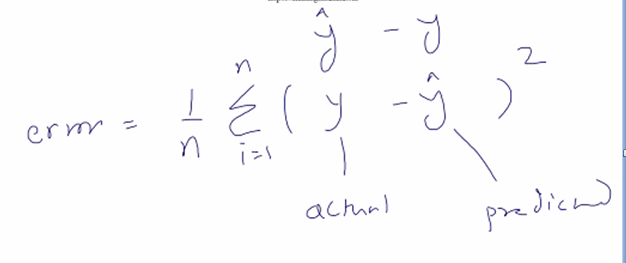
**Small discuss kr lete he:**

Humko nikalna he mean of squared error: [(yhat-y)^2]/no. of records OR points

Is se hume error ka pta lgega , humara goal he is error ko 0 krna

To phir loop me m and c ko change kr k vo input wale x se y’ nikalege and phir usko actual y me se ghata k error.

Ab isko differential se kese kr skte he:



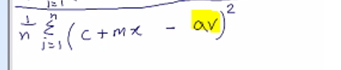
Ab man lo y humara vo vector jo humne input data se nikala he, and y’ vo he to humne m and c ki random value le k x ko le k nikala ,

Hum error nikalege and 0 nahi aayi to phir m and c ko change krege ,

To isko maan to small change krna he

To is ko equation maan lo and iska partial differentia tion kr do ,with respect to m and wrt c.

Like this:



Yaha pr y ko spilt kr diya he equation me and av actual vector jo line.data se nikala he.

Predicted is : c+mx

Actual is : av

Error=predicated-actual;

(c+mx -av)^2

Ab isme c and m ko partially cange krte he , mtlb derivative krte he

To derivative k baad esa bnaega

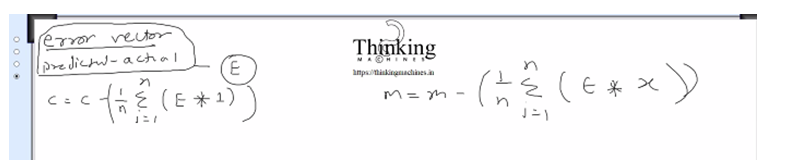
E\*1 and E\*X , and summation of both till n

Ab ye E\*1 c ko change kiye to aaya ab to kitna change kiya usko nikalne k liye ye aa

Ab upar wali equation ka partial derivative krege to esa bnegag,[power rule, constant rule applied]

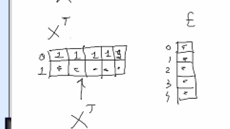
Output ko c me se minus kr dete he , means c ko thoda se upar ya nicke khiskaya

Similarly for m



To isko alaga alag likhne ki bjay m and c ka ek vector le and usko

Ab is equation ko vector ka use kr k solve kre to:

 = 

1\*E , c me jayega=> summation of 1 to n of E\*1

X\*E m me jayega=> summation of 1 to n of E\*X

Ab humko minus krna he to existing model me se is naye model ko minus krege

To partiall khiskega m and c usko dekhne k c- error= to pta chl jayega kitna kishkaya he

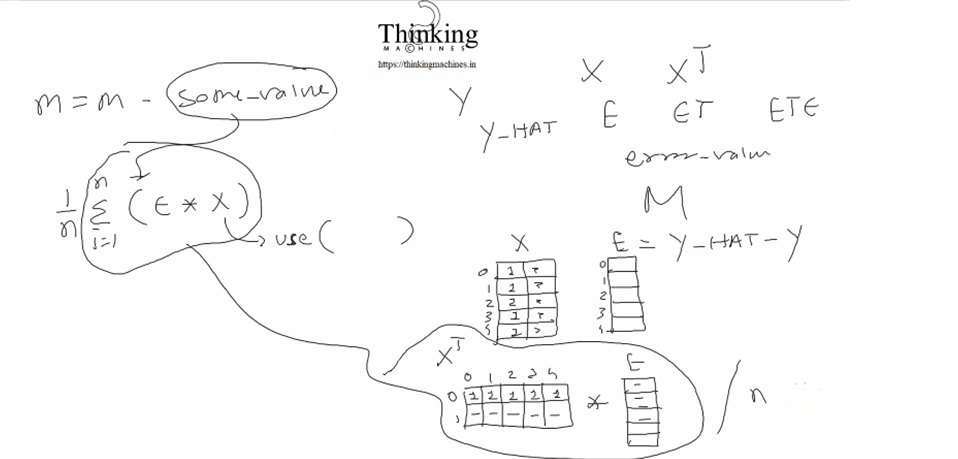
Iske baad bnega linear regression.fit page 54-58 and part 5 ends se. part 5k end tk ye model vector return krne ka kaam,

Linear\_regression\_fit ek function he jo **linear\_regression.c** file me he testcases folder k andr

. Ye in progress he

Ye

Now part 6:



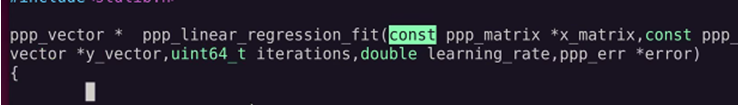
Error vector ko input k sath multiply krege to phir model aa jayega , usko actual model me se minus kr dege.

Old model me kuch value minus ho jayegi and naya model bn jayega.updated wala

Ab nayi value k sath phir se derivative nikalege and so on….

­­­­CODE:

Ab pura main function change, ek function bnega usi file me



Ye input matrix lega and actual vector lega and no. of iteration lega and learning rate lega.

Ab linear regression fit() me loop lagane tk me part 5 ends

Ab loop k andr ka program chle iske liye humko kuch functions ki need he : ye sb part 5 me he

1. sss\_vector\_write\_csv() done, files: vector.h and vector.c
2. sss\_vector\_transpose() Done : vector.c and vector.h me , string.h included for memcpy used
3. sss\_multiply\_vector\_get\_scalar() done math.h me and math.c me
4. sss\_multiply\_vector\_witht\_scalar() done math.h me and math.c me
5. transpose vector me container dalna ,, ye 5 point k baad part 6 me page no. 16tk baat pahuch jayegi

done all 5 points

ss lagana he.

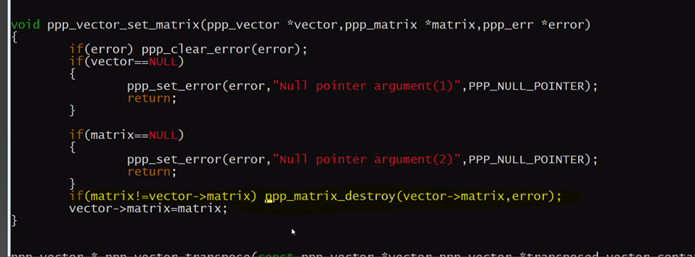
Next: testing me issue tha that is vo user ne same paas kiya

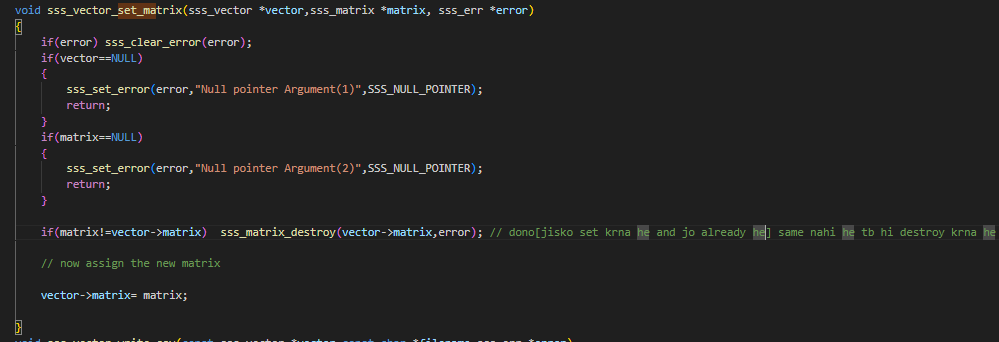
To uska solution humne solution nikala tha math.c me



But actual me solution vector\_set\_matrix me krna he like below

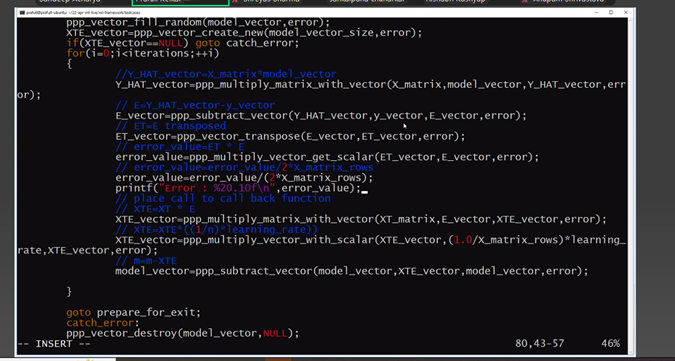
And math . c sir wala code rkhege and if else hatega.: it is done





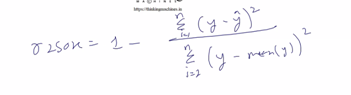
Compiled and libray done:

Ab for loop ka code in linear\_regression.c k main() me



Paart 6 me bakki kaam ye search kr k yaha se start krna heL:

**R2 score k liye sum and mean lagege**



Ye he r2 score ka formula :

To ml\_math.h me function:

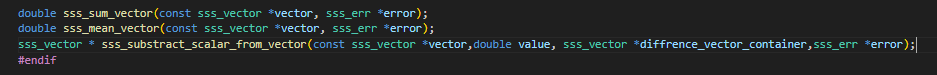


Ml\_math.c me iski definition

Formula hisab se :

Mean nikalne k liye function, summation k liye function, difference k liye function scalar from vector

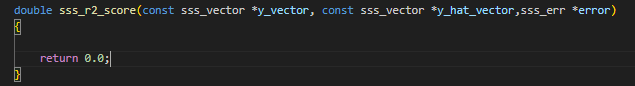
Math.h me declaration:



Math .c me definition

Ye teeno function likh k compile ho gaye he

Ab next : r2score ki defition ml\_math.c me:



Ye likh ke compile ho gaya he

Ab r2 score ka testcse:

Iske liye ek function bnege linear\_regression\_predict()

And ek main()

File in testcase folder: linear\_regression\_predict.c

Ye file bn k compile ho gayi he.

Ab :



Linedata.c generator ka use kr ke line.data bnate he: modeify n=200 in .c file and then compile

 run it

Now split it :

 200 records he



Now copy this trn and tst csv to testcase folder

And compile linear\_regression.c

gcc linear\_regression.c -I ../include -I ../private/include  -I ../include/error -I ../private/include/error  -I ../include/vector  -I ../include/matrix -I ../include/math  -L ../lib -lsss -o linear\_regression.out

since as a lib user we are not suppose to give private path, but since we are using the seterror() , which is in private header file , that’s why we need to give private path for error folder.

Now train the data: using linear\_regression.out and passing training csv and learning rate.



**Note: if we not get the error printing taking long time then chage the learning rate to more small.**

Ye iteration and learning rate k sath change kr kr ke is error tk pahuche.

Ispe model ye aaya:

|  |
| --- |
| C=17830.65 |
| M=-126.51 |

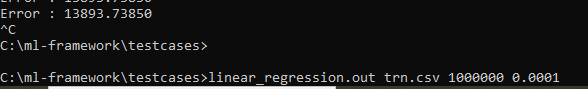
Ab learning rate thoda km kr k dekha to ye aaya



Jo jki phle se jyada he

Learning rate is 0.0001 is fine and iteration is 1lakh

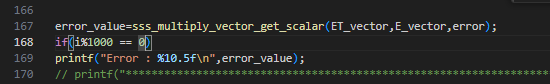
Abhi ek baar isi learning rate k sath 10 lakh iteration kr k dekhte he



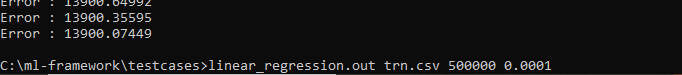
10lakh pr ye aaya 13893 but program print nahi kr raha he , printf ko mod me chalate he

Iteration % 1000

Code change in fit()



Compile and run again with same learn rate 0.0001 but iteration= 5lakh



Ab 10 lakh krte he



Note: printf bahut time leta he , %1000 krne k baad program jldi execute ho gaya.



15 lakh and 10 lakh wale me jyada change nahi aaya

So final is 10 lakh iteration wala , learn rate 0.0001 hi sahi he , dusri learning rate k sath issue aa raha he.

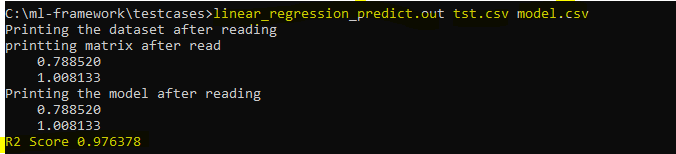
Ye final model c=0. 78851, m= 1. 008133

Ab accuracy check krte he using test data , with predict function.

Lets compile the file:

gcc linear\_regression\_predict.c -I ../include -I ../private/include -I ../include/error -I ../private/include/error -I ../include/vector -I ../include/matrix -I ../include/math -L ../lib -lsss -o linear\_regression\_predict.out

pass testing csv file and model .csv that is created by regression.out programme.



Humari 97% aayi accuracy.is model and test data k liye.

Now

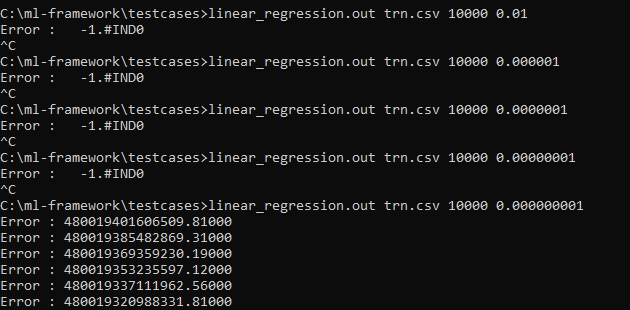
**Go to utility and take another data from kaggle:**

**Take housedata.csv: split in trn.csv and tst.csv**

****

**Ab trn and tst ko utility folder se testcase folder me copy kro.**

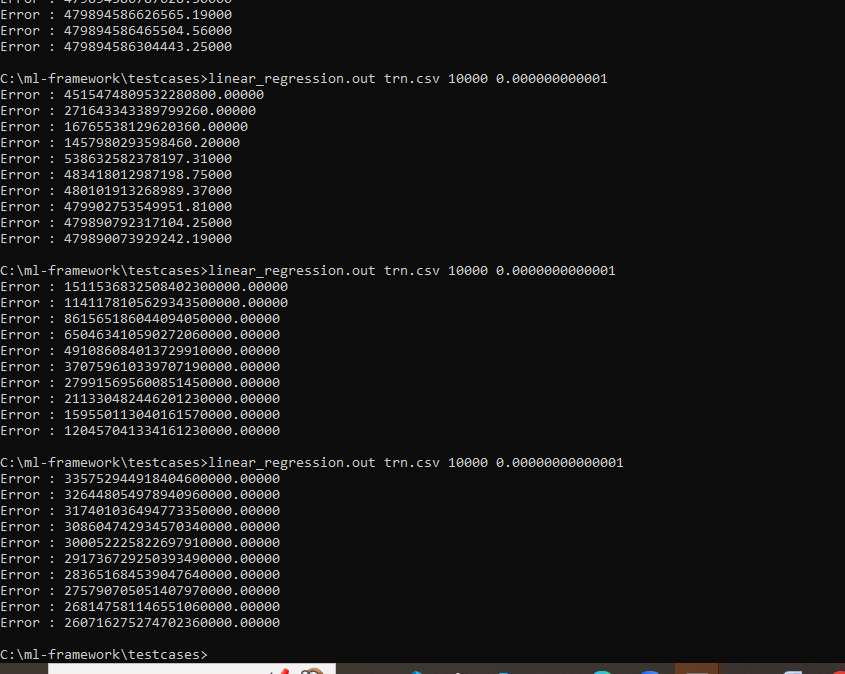
**Now trn se model ko train kro by changing learning rate and iterations:**

****

**Isme learing rate change kr kr ke infinity se no. pr aaye**

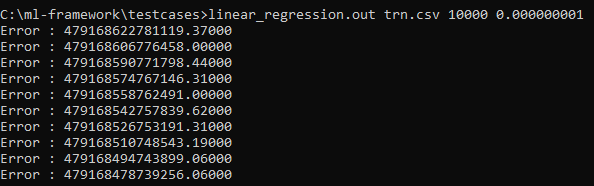
**Note: if we not get the error printing taking long time then chage the learning rate to more small.**

**Ab isko or change kr ke dekho ,**

****

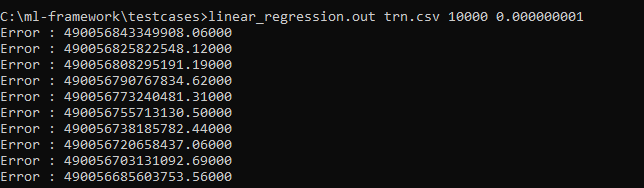
**Humne dekha ek point pr error bdene lagi**

**So final linear\_regression.out trn.csv 10000 0.000000001 ye learning rate sahi he**

****

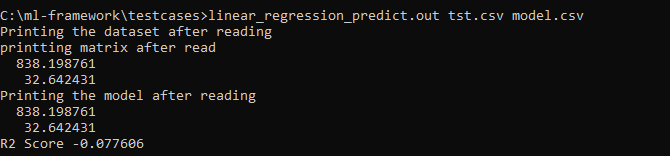
**Ab iterations change kr k dekhte he.**

**Same again run kiya to ye aaya:**

****

**15 digit se km nahi ho raha he.**

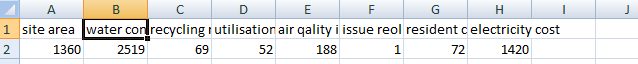
**Lets test the model:**

****

**7% aa rahi he accuracy.**

**Lagta he data linear nahi he kyuki humne actual data k sath chedchad k I thi**

**Abhi actual data lete he jisme multiple dfeature he:**

****

**Jisme 10000 rows he**

****

**Copy trn and tst on testcase folder and train the model by changing learning rate and keeping iteration to 1000 , when get the pick point**

Iteration 1000: change in learning rate:

68448734189285200 0.0000000001

171256341131714720 0.000000001

16681450466505842 0.00000001 \*

17856249257073246

1.#INF0 0.0000001

674408986711671420 0.00000000001

Inme best learing rate he 3rd wali 0.00000001

**Keep the best learning rate** 0.00000001 **constant and try by changing the iteration now**

0.00000001

16681450466505842 10000

1070792981800664 50000

527303812113227 100000

69537339085 1000000

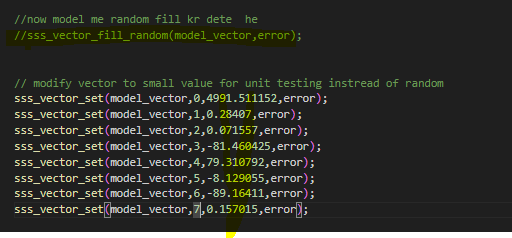
**Check the pick point while changing the iteratin and then save the final model**

|  |
| --- |
| 4991.511 |
| 0.28407 |
| 0.071557 |
| -81.4604 |
| 79.31079 |
| -8.12906 |
| -89.1641 |
| 0.157015 |
|  |

**R2 Score -6.912489**

**Final model: ab ek kaam krte he same learning rate and 100000 iterationk baad same model ko lekr 100000 iterationor krte he.**

**We need to feed this model from code. In linearregression.c**

****

**Compile the code, in testcase folder and run the code with learning rate**

**Ab 10 lakh iteration k model se naya prediction:**

**linear\_regression.out trn.csv 10000 0.00000001**

saved model k bad agle 1lakh iterations se naya :

R2 score R2 Score -2.743534

Error : 32627875734.85931

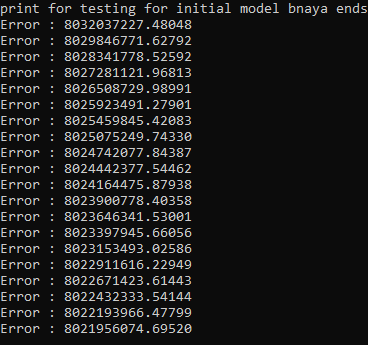
saved model k bad agle 1lakh iterations se naya :

R2 Score 0.028624

Error : 8035286038.12410

**Ab 10 lakh + 10lakh iteration k model se naya prediction:**

**linear\_regression.out trn.csv 10000 0.00000001**

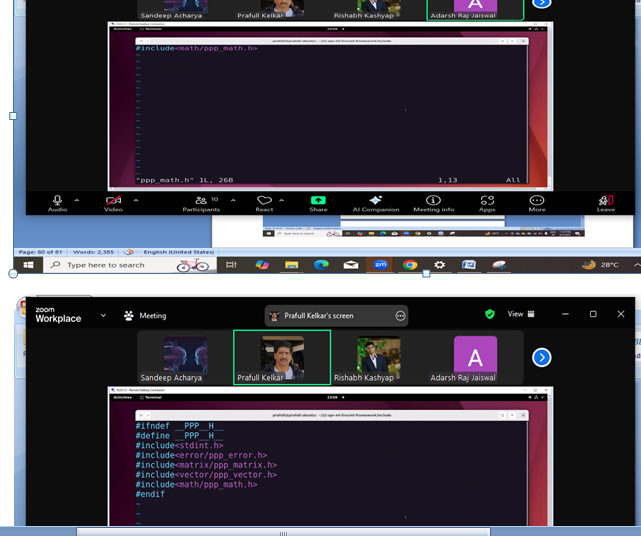


Is baar error jyada deviate nahi ho rahi he means ye model is final with the accuracy

R2 Score 0.029761

Error : 8021956074.69520

compile time pr include krna ka kaam khtm krne wala kaam

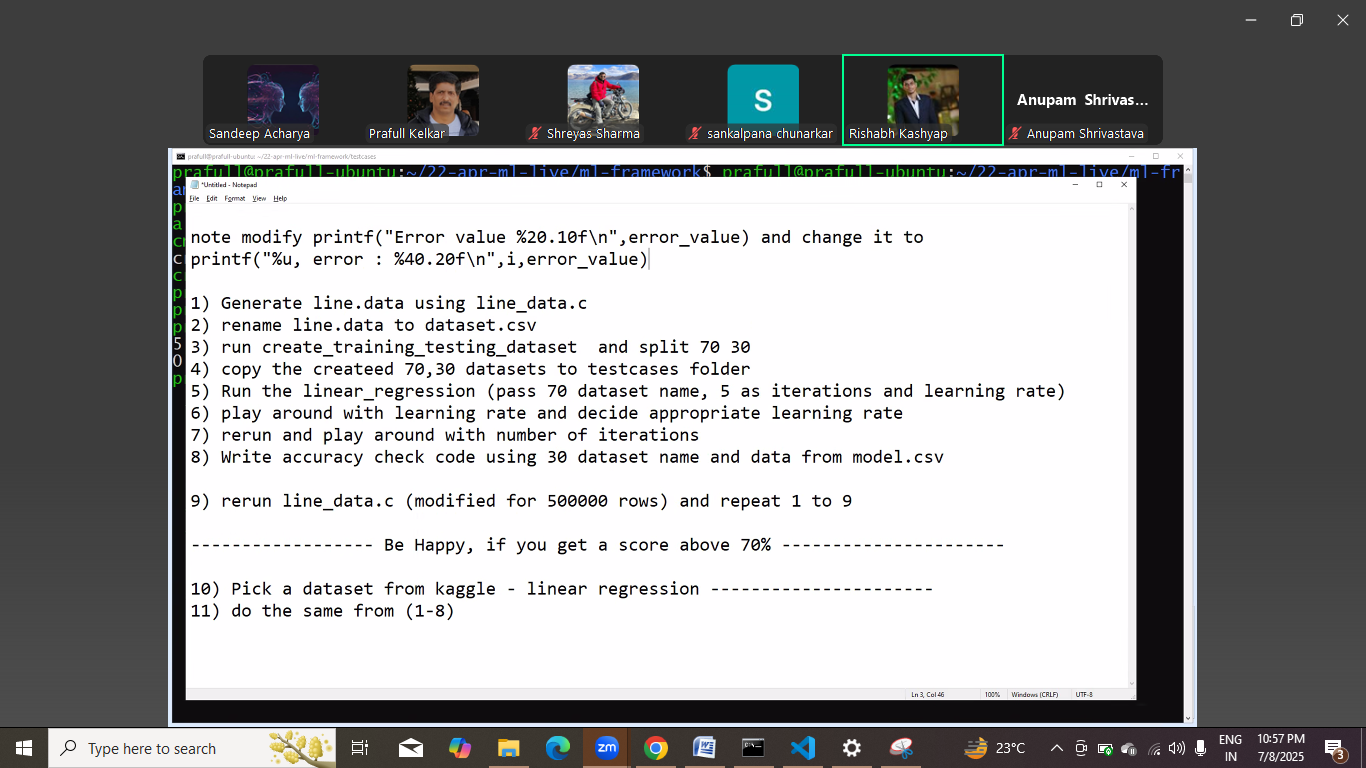


Ek file bnegi include folder me ppp\_math.h same for matrix vecror and all

And usme vo sub folder se file bahar lane ka kaam

come and start: yaha se start krna he part 4 me

**ASSIGNMENT AFTER : fit() and predict()**

****

**This is done**

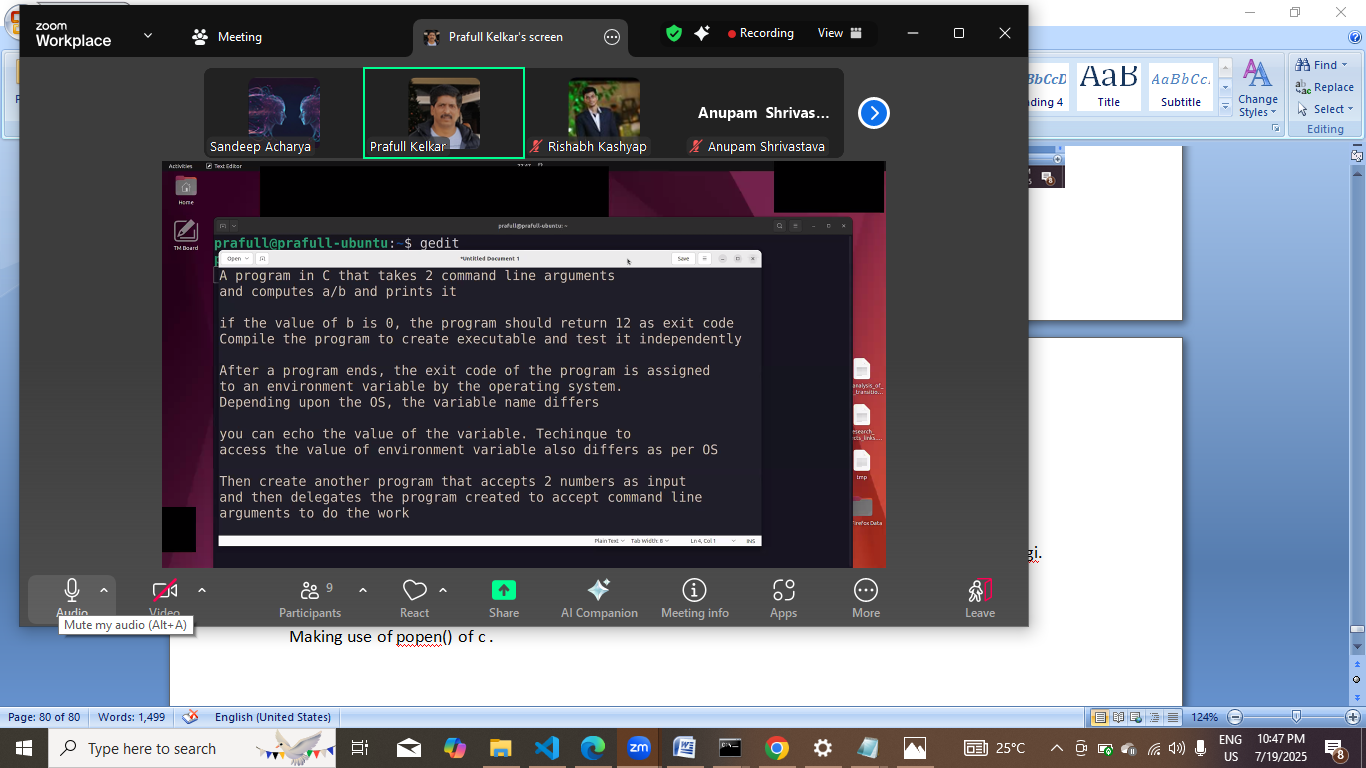
**Below one is pending:**

To ab python me krte he same.

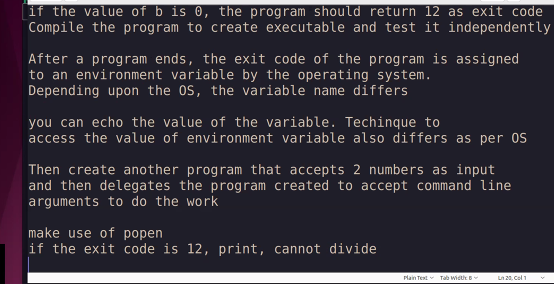
Work : same part python me bnana he ,linear\_regression, without creating library , bni bani mil jayegi.

Determine krna ki time kitna lg raha he same iteration same learning rate, between c and python.

Making use of popen() of c .



Exit code print krwane se phle se dusra program nahi chalana he nahi to uska exitcode write ho jayega.



Is se humko smjh aayega ki kese ek program se dusra program invoke krna and data pass krna.

Fayda: linear regression me gnu plot ko kaam pr lagana he. And showoff

Ek ldke ne link share ki resume me but kisi ne use dekha nahi but jb dekha to fastest hire kr k hire kr liya

Talent.

Vo callback ka use krke

And basically popen() k bare me sikkha he

Now scaling techniques

Final jo fair krna he coding me: part 6 page 40 se pending he:

1. callback wali baat in linear\_regression\_fit
2. R2 score wali baat
3. Linear\_regression\_predict

I thing part 6 tk ka conve ho gaya he ab part se start krna he and

Part 1&2 and 3&4 ko revise krna he.