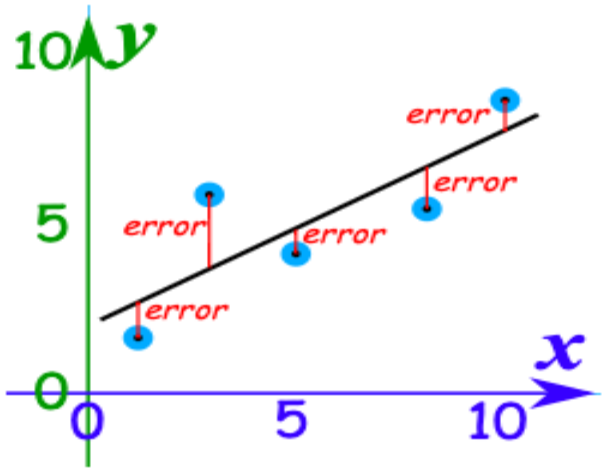
VLSI PROJECT
LEAST SQUARES REGRESSION

DEFINITION :

The Least Squares Regression Line is the line that makes the vertical distance from the data points to the regression line as small as possible. It's called a "least squares" because the best line of fit is one that minimizes the variance (the sum of squares of the errors). The aim to find the equation that fits the points as closely as possible**.**

HOW DOES LSR WORK?

It works by making the total of the square of the errors as small as possible (that is why it is called "least squares"):



The straight line minimizes the sum of squared error. So, when we square each of those errors and add them all up, the total is as small as possible.

THE LINE

Our aim is to calculate the values m (slope) and b (y-intercept) in the equation of a line :

$y = mx + b$

Where:

y = how far up
x = how far along
m = Slope or Gradient (how steep the line is)
b = the Y Intercept (where the line crosses the Y axis)

STEPS

To find the line of best fit for N points:

Step 1: For each (x,y) point calculate x2 and xy

Step 2: Sum all x, y, x2 and xy, which gives us Σx, Σy, Σx2 and Σxy (Σ means "sum up")

Step 3: Calculate Slope m:

$m = (N\Sigma(xy) - \Sigma x\, \Sigma y) / (N\, \Sigma(x2) - (\Sigma x)^2)$

(N is the number of points.)

Step 4: Calculate Intercept b:

$b = (\Sigma y - m\, \Sigma x) / N$

Step 5: Assemble the equation of a line

$y = mx + b$

EXAMPLE

Example: SAINTO have data of how many hours of sunshine vs how many ice creams were sold at the shop from Monday to Friday:

"X" Hours of Sunshine	"Y" Ice Creams Sold
2	4
3	5
5	7
7	10
9	15

find the best m (slope) and b (y-intercept) that suits that data

$y = mx + b$

Step 1: For each (x,y) calculate x2 and xy:

x	y	x2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135

Step 2: Sum x, y, x2 and xy (gives us Σx, Σy, Σx2 and Σxy):

X	y	x2	xy
2	4	4	8
3	5	9	15
5	7	25	35
7	10	49	70
9	15	81	135
Σx: 26	Σy: 41	Σx2: 168	Σxy: 263

Also N (number of data values) = 5

Step 3: Calculate Slope m:

m = (N Σ(xy) − Σx Σy) / (N Σ(x2) − (Σx)^2)

= 5 x 263 − 26 x 415 x 168 − 262

= 1315 − 1066840 − 676

= 249164 = 1.5183...

Step 4: Calculate Intercept b:

b = Σy − m ΣxN

= 41 − 1.5183 x 265

= 0.3049...

Step 5: Assemble the equation of a line:

y = mx + b

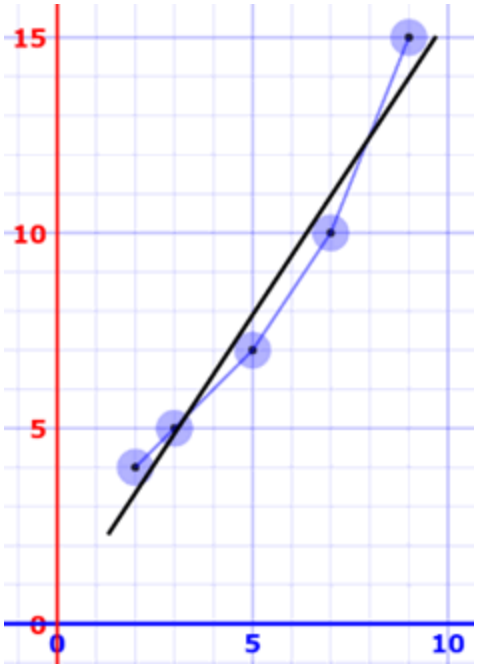
y = 1.518x + 0.305

Let's see how it works out:

x	y	y = 1.518x + 0.305	error
2	4	3.34	−0.66
3	5	4.86	−0.14
5	7	7.890	.89
7	10	10.93	0.93
9	15	13.97	−1.03

Here are the (x,y) points and the line y = 1.518x + 0.305 on a graph:

least squares example graph



SAINTO hears the weather forecast which says "we expect 8 hours of sun tomorrow", so he uses the above equation to estimate that he will sell
y = 1.518 x 8 + 0.305 = 12.45 Ice Creams

PYTHON AND VERILOG IMPLEMENTATION OF LSR

The steps discussed above have been implemented in python. Wherever there are operators such as +, -, x instead of them verilog codes have been used to perform those operations. In verilog test bench file value of variable ‘a’ and ‘b’ are written and then it returns the appropriate result which is captured and being used for calculations further on.

INPUT FOR PYTHON IMPLEMENTATION
[2.01357, 3.1251, 5.0, 7.179, 9.012]
[4.01, 5.038, 7.012378, 10.0, 15.08]

OUTPUT