



Problem Sheet 05

Problem 1: Linear Regression with Error in Both Variables

Consider the following data collected from an experiment:

x	0	2	4	6	9	11	12	15	17	19
y	5	6	7	6	9	8	7	10	12	12

Obtain a linear fit of the form $y = a_0 + a_1x$. Compute the correlation coefficient, R as well.

Next, repeat the problem with dependent and independent variables switched. In other words, obtain the linear fit $x = b_0 + b_1y$. Also compute the correlation coefficient, R.

From these values of b_0 and b_1 , compute the corresponding values a_0 and a_1 .

Compare the two fits by plotting the straight line fits as well as the original data on the same plot. Use the Scilab command `plot(x, y, ' .b')` to plot the original data as dots.

Problem 2: Multi-Linear Regression

Two different methods were shown in the lectures for linear regression of the form

$y = a_0 + a_1x + a_2w$. The first method was to minimize the sum of square errors,

$$I = \sum_{i=1}^n (y_i - (a_0 + a_1x + a_2w))^2 \text{ by equating } \frac{\partial I}{\partial a_0} = \frac{\partial I}{\partial a_1} = \frac{\partial I}{\partial a_2} = 0 \text{ and solving for } a_0, a_1, \text{ and } a_2.$$

How are the values of a_0 , a_1 , and a_2 using this method related to the data x_i , w_i , and y_i ?

An alternative method was the matrix method, where we created data matrices to represent the data in the form $\mathbf{Y} = \mathbf{A}\boldsymbol{\phi} + \mathbf{E}$, and obtained the solution as $\boldsymbol{\phi} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}$. Show that for the above problem involving one dependent and two independent variables, the matrix method leads to the same result as that obtained using the first method.

For the following data, obtain the straight-line fit of the form $y = a_0 + a_1x + a_2w$ using both the methods described above.

x	4	5	6	7	11	13	18	19
w	0.4	0.8	0.85	0.95	1	1.25	1.55	1.85
y	4.5	5.9	6.25	6.75	7.7	8.85	10.75	11.85

Problem 3: Nonlinear Regression

Consider the Antoine equation that relates vapour pressure of a pure fluid to its temperature:



$$\ln(p_{\text{sat}}) = a - \frac{b}{T + c}$$

where, a , b , and c are the Antoine parameters. The following data is available for acetone:

T(K)	259.2	273.4	290.1	320.5	390.3	446.4	470.6	508.1
p_{sat}	0.04267	0.09497	0.21525	0.74449	5.655	17.682	26.628	47.0

Use an appropriate nonlinear parameter estimation method to obtain the coefficients a , b and c .

An additional experiment was conducted and the saturation pressure at 350.9 K was 2.0157 bar.

How well does the model predict this additional data?

Problem 4: Reaction Kinetics using Linear Regression

The rate of an enzymatic reaction is given by the following expression:

$$r = \frac{k[S]}{K_m + [S]}$$

The problem of estimating k and K_m can be converted to linear regression by inverting the above expression and defining:

$$x = \frac{1}{[S]}, \quad y = \frac{1}{r}$$

The following data was obtained in the lab:

[S]	1.233	0.540	0.442	0.258	0.198	0.162	0.130	0.128
r	5.970	3.319	2.253	2.547	1.493	1.182	1.095	0.869

1. Obtain $y = a_0 + a_1x$ and hence find the values of k and K_m .
2. Now, take x as the dependent variable and obtain $x = b_0 + b_1y$ and hence find the values of k and K_m .
3. Compare the two values and explain your results.

Problem 5: Polynomial Regression

The following data has been obtained using a general nonlinear function:

x	-5	-3	1	2	4	7	8	9
y	15	1	7	9	34	68	83	119

Since we have a total of 8 data points, we can fit a polynomial upto order 7. For this problem, fit polynomials of all orders from 1 to 7. Compute the correlation coefficient for each of them.

Comment on what order polynomial would best represent the data?



Plot the data given in the above table as dots. Additionally, plot the polynomial approximations for the first-order, seventh-order and best-fit polynomials as three lines. Compare and comment.

Two more “test” data points were collected: (5, 42) and (10, 132). How well do the above polynomials predict the values of \hat{y} for these two data points?

Problem 6: Interpolation methods

For the data given in Problem 5, using any two interpolating methods taught in the video lectures. Plot the given data (as dots) as well as the interpolating function (as lines).

How well do the two methods predict the values of \hat{y} for the two test data points?