Lab Session 6

MA423 : Matrix Computations Lab 2023 R. Alam

Least squares problem

1. Over 400 years ago Galileo, attempting to find a mathematical description of falling bodies, studied the paths of projectiles. One experiment consisted of rolling a ball down a grooved ramp inclined at a fixed angle to the horizontal, starting the ball at a fixed height h above a table of height 0.778 meters. When the ball left the end of the ramp, it rolled for a short distance along the table and then descended to the floor. Galileo altered the release height h and measured the horizontal distance d that the ball traveled before landing on the floor. The table below shows data from Galileos notes (with measurements converted from puntos to meters)

   Data from Galileos inclined plane experiment.

h	0.282	0.564	0.752	0.948
d	0.752	1.102	1.248	1.410

   Determine a a least squares linear fit $y = c_1 + c_2 x$ with the following MATLAB code:

   ```
   %%% Set up data
   h = [0.282; 0.564; 0.752; 0.940]; d = [0.752; 1.102; 1.248; 1.410];

   %%% Form the 4x2 matrix A and solve for the coefficient vector c.
   A = [ones(size(h)), h]; c = A\d;   % solution of the LSP Ac =d
   cc = flipud(c);    % order the coefficients in descending
   % order for polyval
   %%% Plot the data points
   plot(h,d,'b '), title('Least Squares Linear Fit'), hold
   xlabel('release height'), ylabel('horizontal distance')

   %%% Plot the line of best fit
   hmin = min(h); hmax = max(h); h1 = [hmin:(hmax-hmin)/100:hmax];
   plot(h1, polyval(cc, h1), 'r'), axis tight
   ```

   This code produces the line y = 0.4982 + 0.9926x. This line is plotted along with the data points. A measure of the amount by which the line fails to hit the data points is the residual norm $r := \|Ac - d\|_2$. Determine r.

2. Solve the following least-squires problems (linear regression) using the method of normal equation. For each problem, plot the data and their best linear fit in a single plot.

Exercises

5.5.1. Find the straight line $y = \alpha + \beta\, t$ that best fits the following data in the least squares sense: (a)

t_i	-2	0	1	3
y_i	0	1	2	5

(b)

t_i	1	2	3	4	5
y_i	1	0	-2	-3	-3

(c)

t_i	-2	-1	0	1	2
y_i	-5	-3	-2	0	3

5.5.2. The proprietor of an internet travel company compiled the following data relating the annual profit of the firm to its annual advertising expenditure (both measured in thousands of dollars):

Annual advertising expenditure	12	14	17	21	26	30
Annual profit	60	70	90	100	100	120

(a) Determine the equation of the least squares line. (b) Plot the data and the least squares line. (c) Estimate the profit when the annual advertising budget is \$50,000. (d) What about a \$100,000 budget?

5.5.3. The median price (in thousands of dollars) of existing homes in a certain metropolitan area from 1989 to 1999 was:

year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
price	86.4	89.8	92.8	96.0	99.6	103.1	106.3	109.5	113.3	120.0	129.5

(a) Find an equation of the least squares line for these data. (b) Estimate the median price of a house in the year 2005, and the year 2010, assuming that the trend continues.

5.5.4. A 20-pound turkey that is at the room temperature of 72° is placed in the oven at 1:00 pm. The temperature of the turkey is observed in 20 minute intervals to be $79^\circ, 88^\circ$, and 96°. A turkey is cooked when its temperature reaches 165°. How much longer do you need to wait until the turkey is done?

♡ 5.5.5. The amount of waste (in millions of tons a day) generated in a certain city from 1960 to 1995 was

year	1960	1965	1970	1975	1980	1985	1990	1995
amount	86	99.8	115.8	125	132.6	143.1	156.3	169.5

(a) Find the equation for the least squares line that best fits these data.
(b) Use the result to estimate the amount of waste in the year 2000, and in the year 2005.
(c) Redo your calculations using an exponential growth model $y = c\, e^{\alpha\, t}$.
(d) Which model do you think most accurately reflects the data? Why?

5.5.6. The amount of radium-224 in a sample was measured at the indicated times.

time in days	0	1	2	3	4	5	6	7
mg	100	82.7	68.3	56.5	46.7	38.6	31.9	26.4

(a) Estimate how much radium will be left after 10 days.
(b) If the sample is considered to be safe when the amount of radium is less than .01 mg, estimate how long the sample needs to be stored before it can be safely disposed of.

5.5.7. The following table gives the population of the United States for the years 1900-2000.

year	1900	1920	1940	1960	1980	2000
population – in millions	76	106	132	181	227	282

(a) Use an exponential growth model of the form $y = ce^{at}$ to predict the population in 2020, 2050, and 3000. (b) The actual population for the year 2020 has recently been estimated to be 334 million. How does this affect your predictions for 2050 and 3000?