CHEN 2450

HOMEWORK 6

LINEAR LEAST-SQUARES REGRESSION

Submit all your homework using Jupyter notebooks. You are expected to include proper text and discussion for each and every problem including appropriate headings and formatting. You will lose points if your reports are not readable or just include code with a few print out statements. No exceptions.

Problem 1 (30 pts)

We saw in class how to derive the regression equations for a straight line and for a second order polynomial. Here, you will derive the equations for a polynomial of degree 3. Consider N data points x_i and y_i. We wish to fit a model that has the form:

$$y = a_0 + a_2 x^2 + a_3 x^3 \tag{1}$$

1. (10 pts) Write the least squares error $S = \sum (y_i - f_i)^2$ and identify the unknown quantities.
2. (10 pts) What are the conditions for which S is a minimum?
3. (10 pts) In light of the previous question, write the system of linear equations that must be solved to find the unknown quantities.

Problem 2 (30 Pts)

The heat capacity C_p is the number of heat units needed to raise the temperature of 1 kg of matter by one degree. The heat capacity of a gas at various temperatures T is given by the following data:

C_p (J/Kg/K)	-50	-30	0	60	90	110
T (K)	1270	1280	1350	1480	1580	1700

Table 1: Heat capacity c of a gas as a function of absolute temperature, T

1. (10 pts) Perform a straight line fit for the data in 1 and plot your fit on top of the original data. Comment on your results. It is best to show the input data as dots on the plot to distinguish between your fit and the original data.

2. (10 pts) Perform a quadratic fit for the data in 1 and plot your quadratic fit on top of the original data as well as the straight line fit. Comment on your results. You can either use the Normal equations or the standard form to fit the data.

3. (10 pts) Based on the plots, which fit do you expect to be a more accurate model of the data? What number can quantify which curve is a better fit? Compute that number and compare it for both fits.

Problem 3 (40 Pts)

The growth rate k (# bacteria per day), of a certain bacteria is reported as a function of oxygen concentration c (mg/L)

c (mg/L)	0.5	0.8	1.5	2.5	4
k (#/day)	1.1	2.4	5.3	7.6	8.0

Table 2: Bacertial growth rate k (per day) as a function of oxygen concentration c(mg/L)

It is known that such data can be modeled by the following equation:

$$k = \frac{k_{\max} c^2}{c_s + c^2} \tag{2}$$

where c_s and $k_{\max}$ are to be determined. We plan on using linear least-squares regression to fit this model to the data in 2.

1. (10 pts) Use algebraic manipulation to transform this equation into a form suitable for regression to a straight line of the form: $y = a_0 + a_1 x$. (HINT: what happens if you take the reciprocal, $1/k$, of the model equation 2?)

2. (10 pts) Write down the linear system of equations that you need to solve to find a_0 and a_1. Use either the normal equations or the traditional form.

3. (5 pts) Solve the above system of equations for a_0 and a_1.

4. (10 pts) Compute the R^2 value for this fit and comment on your finding.

5. (5 pts) For an oxygen concentraion of 2.1 mg/L, what is the bacterial growth rate?