Numerical Analysis

Worksheet 18

SUMMARY Data Linearization and Least-Squares Curve Fitting
READING Recktenwald, Sec 9.1, pp 455–468; Mathews & Fink, p 263–278

Fitting Data to Nonlinear Functions: OR Making Nonlinear Relationships Appear Linear
Isn't there some way we could transform the equation $y = be^{ax}$ and $y = bx^a$ so that a linear relationship would appear? Then we could use our previously defined normal equations. This process is called **data linearization**.

Think about introducing some new variables Y and X such that there is a linear relationship between Y and X even though y and x are non-linearly related.

A Harder One
How could you pick Y and X so that you could solve the normal equations and fit data to $y = \alpha x e^{\beta x}$?

Exercise
What transformations would you have to make to $y = \dfrac{x}{ax+b}$ to produce a linearized relationship between Y and X

Linear Algebra Approach To Discrete Least Squares Aproximation

What is the system of equations we are trying to solve when finding a line of best fit $y = ax+b$ for m data points (x_k, y_k)?

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \\ \vdots &= \vdots \\ ax_m + b &= y_m \end{aligned}$$

1. What are the known parameters here and what are the unknown variables ?

2. Write the system as a linear system $A\vec{c} = \vec{y}$

3. What are the dimensions of the matrix A and the vectors $\vec{c}$ and $\vec{y}$?

4. It turns out that when one can not solve the linear system $A\vec{c} = \vec{y}$ from Linear Algebra we know that instead you should solve the system $A^T A\vec{c} = A^T\vec{y}$. The vector $\vec{c}$ while not in the column space of A will have the minimum distance from vectors lying in the column space. This equation $A^T A\vec{c} = A^T\vec{y}$ is known as the Matrix Form of the Normal Equations. How many equations in how many unknown does it represent?

Least Squares Fit Using Other Functions

There is no reason why one has to pick the linear function $P(x) = ax + b$ to fit the data to. That is, one does not have to do the fit to a linear function (or even transform the data so that a linear relationship can be established between input and output variables).

Suppose

$$y = P(x) = c_1 f_1(x) + c_2 f_2(x) + \ldots + c_k f_k(x) = \sum_{i=1}^{k} c_i f_i(x)$$

Thus the best fit problem will become one of finding the k c_i values which cause the least square error to be minimized.

NOTE If k, the number of basis functions, is equal to the number of data points, m, then one is really doing **interpolation** instead of polynomial best fit with the basis function $\{f_i(x)\}_{k=1}^{m}$. Generally these functions would be $m-1$-degree polynomials to reduce the situation to the Polynomial Interpolation problem we have previously studied.

Exercise

5. Write down the expression $E(\vec{c}) = E(c_1, c_2, \ldots, c_n)$ which must be minimised:

Polynomial Of Best Fit

To simplify things, let's assume that the basis functions are monomials and we have m data points

$$P(x) = \sum_{k=0}^{n} a_k x^k, \qquad \text{where } n + 1 < m$$

6. Why does $n + 1$ have to be less than m?

EXAMPLE

Let's find the normal equations whose solution is the a_k values. (HINT: write it in Matrix Form.)

The above represents the $n + 1$ **normal equations** in the $n + 1$ unknowns $a_0, a_1, \ldots a_n$

Using `polyfit` in MATLAB

Let's try and find a second degree polynomial which fits the data $(0, 1)$, $(0.25, 1.2840)$, $(0.5, 1.6487)$, $(0.75, 2.1170)$ and $(1.00, 2.7183)$

7. What will the dimensions of the matrix equation to be solved look like? (i.e. what is n and what is m?)

8. Find the coeffecients of the polynomial using `polyfit`
(Write down what commands you use)

5. Plot the data and the polynomial of best fit on the same graph.
(Write down what commands you use)