CS685A: Data Mining

Endsem: Part 2
Total Marks: 50

13th December, 2020: 10:45–11:30 am

Study the table of numbers the link of which has been sent to you by email. (Please notify immediately if you have not received the email or you cannot download the csv file.)

The value y is a *sum* of three functions f_i of x_i:

$$y = f_1(x_1) + f_2(x_2) + f_3(x_3)$$

The values of x_i are between 0.05 and 1.00 and can vary in intervals of 0.05 only.
The function f_i is described as $f_i = c_i \cdot g(x_i)$. Thus,

$$y = c_1 \cdot g_1(x_1) + c_2 \cdot g_2(x_2) + c_3 \cdot g_3(x_3)$$

The function g_i is either *logarithmic* or a *polynomial* of x_i.
The logarithmic form is $\log_2(x_i + l)$ where l is one of $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.
The polynomial form is x_i^d where $d \neq 0$ is one of $\{-5, -4, -3, -2, -1, +1, +2, +3, +4, +5\}$.
The coefficients $c_i \neq 0$ are integers between -500 and $+500$ (inclusive).

Find out the *exact functions* f_1, f_2, and f_3, i.e., the functional forms g_i with appropriate parameters, and the corresponding coefficients c_i.

It is given that there is *at most one* logarithmic function.
Examples of possible functions are:

$$y = 50 \cdot x_1^{+2} + 100 \cdot x_2^{+1} - 55 \cdot \log_2(x_3 + 0.5)$$
$$y = 60 \cdot x_1^{-4} + 25 \cdot \log_2(x_2 + 0.7) - 305 \cdot x_3^{-3}$$

If you cannot figure out all the three functions and their coefficients, write about whatever functions and coefficients you can decipher.
In case you cannot decipher the functions, write about some properties and trends.

Describe your working *in detail*.