Smooth Curve Fitting

Curve fitting is the process of constructing a curve, or mathematical function (e.g. polynomial equation) that has the best fit to a series of data points, possibly subject to constraints. In smooth curve fitting, function is constructed to ***approximately*** fit the data.

In this assignment, we are given set of points and would like to fit a curve using a polynomial equation to them. Using genetic algorithms, find the best **coefficients** that would make the distance between the polynomial and the points **minimum**.

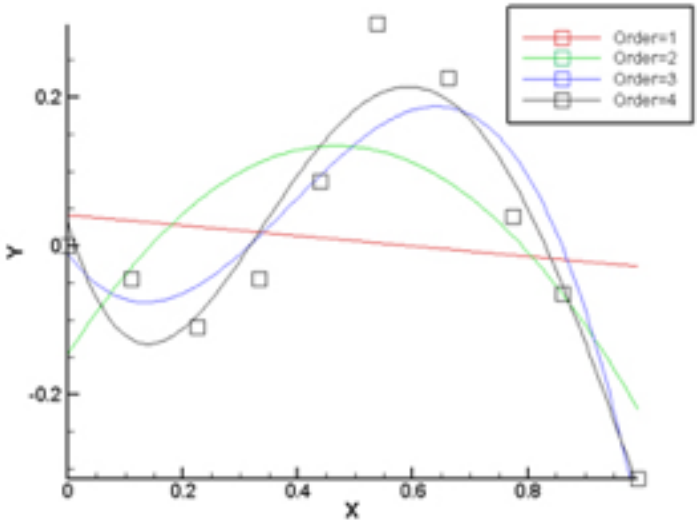

Figure 1: We are given some points, and have 4 different polynomials (of different orders) trying to fit curve to the points

http://en.wikipedia.org/wiki/Curve_fitting

Assume we will fit the data using a 3rd degree polynomial:

$$F(X) = y = a_0 + a_1 x^1 + a_2 x^2 + a_3 x^3$$

Find the coefficients (a0, a1, a2, a3) using genetic algorithm according to the following constrains:

1. $a_i \in [-10, 10]$.
2. The chromosome is represented in a floating point representation and its size equals #coefficients.
3. **Fitness** function is the mean square error. One set of points to try:
   a. P1 = (2 , 8)
   b. P2 = (3 , 27)
   c. P3 = (4 , 64)

4. Best individual is the one with the smallest fitness function because we want to minimize MSE
5. Use non-uniform mutation.

To calculate the fitness (the mean square error) for a given chromosome:

1. For each point p, substitute its X in the polynomial equation to get y_{calc} .
    a. E.g. in the given 3 points, our X's are (2, 3, 4)
2. Then calculate the mean square error:
    a. $\boldsymbol{Error=1/N\ \Sigma(y_{calc} - y_{actual})^2}$, where N = no of points tested, y_{calc} = the y you calculated in the previous step and y_{actual} = is y in the points (P1 , P2 , P3) which corresponds to the X you used to calculate y_{calc} (e.g. 8, 27, 64)
3. P.S. The most optimal coefficients for these 3 points are ($a_0 = a_1 = a_2 = 0$ and $a_3 = 1$), then given X = 4 —> F(4) = 0+0+0+4^3 = 64. For these coefficients Error is 0

Assignment

Given a file of M data sets, for each set, print list of coefficients and the total error. See attached input file.

Input File Structure:

First line, M, represents number of sets.

Each set consist of Line N D, where N is number of points and D is the requested polynomial degree. Then N lines each one representing an (x, y) Point. For example:

4 2 -> 4 points, requested degree is 2 (a0, a1, a2)

1 5 -> one of (x, y) = (1, 5)

2 8

3 13

4 20

Output File Structure:

Consists of M lines, each line has D+1 number for coefficients followed by Error = Total Error. For example, for the above case:

1, 0, 4. Error = 2.1563

Note, above data is not correct.

Number of team members: 3 to 4.

Assignment will be delivered and discussed in next week's labs.