Computational Intelligence Homework Assignment

As discussed in class there are two problems. Your solution will be a description of CI systems used to provide the required prediction. No programming is required, just a description of the representation, evaluation, variation, selection and initialization techniques you would use.

Gestation Period Problem

A data set has been provided to you with lots of information about mothers. Your job is to design a CI system to predict the gestation of the baby. This will help to determine the factors that cause premature births.

The data set has the following information for each individual,

- id - identification number

- plurality - 5= single fetus

- outcome - 1= live birth that survived at least 28 days

- date - birth date where 1096=January1,1961

- gestation - length of gestation in days

- sex - infant's sex 1=male 2=female 9=unknown

- wt - birth weight in ounces (999 unknown)

- parity - total number of previous pregnancies including fetal deaths and still births, 99=unknown

- race - mother's race 0-5=white 6=mex 7=black 8=asian 9=mixed 99=unknown

- age - mother's age in years at termination of pregnancy, 99=unknown

- ed - mother's education 0= less than 8th grade, 1 = 8th -12th grade - did not graduate, 2= HS graduate– no other schooling , 3= HS+trade, 4=HS+some college 5= College graduate, 6= Trade school, 7= HS unclear, 9=unknown

- ht - mother's height in inches to the last completed inch 99=unknown

- wt - mother pregnancy wt in pounds, 999=unknown

- drace - father's race, coding same as mother's race.

- dage - father's age, coding same as mother's age.

- ded - father's education, coding same as mother's education.

- dht - father's height, coding same as for mother's height

- dwt - father's weight coding same as for mother's weight

- marital 1=married, 2= legally separated, 3= divorced, 4=widowed, 5=never married

- inc - family yearly income in $2500 increments 0 = under 2500, 1=2500-4999, ..., 8= 12,500-14,999, 9=15000+, 98=unknown, 99=not asked

- smoke - does mother smoke? 0=never, 1= smokes now, 2=until current pregnancy, 3=once did, not now, 9=unknown

- time - If mother quit, how long ago? 0=never smoked, 1=still smokes, 2=during current preg, 3=within 1 yr, 4= 1 to 2 years ago, 5= 2 to 3 yr ago, 6= 3 to 4 yrs ago, 7=5 to 9yrs ago, 8=10+yrs ago, 9=quit and don't know, 98=unknown, 99=not asked

- number - number of cigs smoked per day for past and current smokers 0=never, 1=1-4 2=5-9, 3=10-14, 4=15-19, 5=20-29, 6=30-39, 7=40-60, 8=60+, 9=smoke but don't know, 98=unknown, 99=not asked

Your job is to describe a computational intelligence prediction system using this data set to predict gestation period. This description should consist of 5 sections, as we described in class,

- Representation

- Evaluation

- Variation

- Selection

- Initialization.

Feel free to use the GP paradigm (since that's what we spent some time on in class) but you may use GA or the general evolutionary computation technique. If you chose to use GP then the representation sub problem becomes one of defining the functions and terminals; variation is somewhat predetermined as crossover or mutation.

Bit Prediction Problem

You are given 10,000 bits. Your job is to predict the next 10,000 bits. The assumption is that there is some algorithm (other than random) that is generating these bits and if you can discover it (using CI) you can predict the next 10,000 and become famous.

As in the previous problem you may use any CI paradigm to design your predictor, and describe it using the standard 5 elements,

- Representation

- Evaluation

- Variation

- Selection

- Initialization.