Discovering Patterns Perfectly Matching the Data

Cristian Vava, PhD*

•CEO of Innovatorium Technologies Corporation, Your Special Analytics Department

Scope

- In science and business we often need to find a formula (also called pattern) describing the relation between a dependent variable and a set of independent variables
- Once found, such formula could be very useful in describing the relation between its input and output parameters
- The main purpose of this white paper is to show a simple method to find such formula using linear algebra
- We'll show that with few (X, Y) pairs it is always very easy to find a formula capable to perfectly match the given data
- As example we'll use the (X, Y) pairs from table 1.

Table 1

X	Y
1	1
2	3
3	5
4	7

Formalizing the Pattern Design

When trying to find a mathematical pattern what we look for is a function F which generates the values of Y from the values of X in such a way as they match the given combinations (X_1,Y_1; X_2,Y_2; X_3,Y_3; X_4,Y_4).

Mathematically the relation can be written as:

$$Y_1 = F(X_1);\ Y_2 = F(X_2);\ Y_3 = F(X_3);\ Y_4 = F(X_4)$$

There are numerous ways to design the function F however, in general for few pairs of data points the easiest and safest way is to use linear algebra.

PATTERN DESIGN ALGORITHM

Generating Matrix

1. Select a set of independent functions (called base).

   In the examples below we'll use polynomial functions:

   [A]: $X^0, X^1, X^2, X^3, \dots$

   [B]: $X^0, X^2, X^3, X^4, \dots$

   [C]: $X^0, X^3, X^4, X^5, \dots$

   Many other functions can be used, common examples are based on but not limited to the exponential and trigonometric formulas.

2. Create the generating matrix M using a number of functions equal to the number of original X,Y pairs.

   For the polynomial functions in example [A]:

$$M = \begin{bmatrix} X_1^0 & X_1^1 & X_1^2 & X_1^3 \\ X_2^0 & X_2^1 & X_2^2 & X_2^3 \\ X_3^0 & X_3^1 & X_3^2 & X_3^3 \\ X_4^0 & X_4^1 & X_4^2 & X_4^3 \end{bmatrix}$$

Where X_1, X_2, X_3, X_4 are the four given values of X. In our approach the independent variables X are embedded in the generating matrix M.

A Little Algebra

3. Create the output vector as:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}$$

4. Compute the coefficients [C]:

Here $[]^{-1}$ represents the inverse of the generating matrix M. In our algorithm M is a square matrix so to be invertible it's determinant must be different from zero.

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} X_1^0 & X_1^1 & X_1^2 & X_1^3 \\ X_2^0 & X_2^1 & X_2^2 & X_2^3 \\ X_3^0 & X_3^1 & X_3^2 & X_3^3 \\ X_4^0 & X_4^1 & X_4^2 & X_4^3 \end{bmatrix}^{-1} * \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}$$

5. The pattern generating formula becomes:

Where M is recreated for any set of X and C is from now on a constant vector.

$$Y = M * C$$

EXAMPLES

Example: Excel implementation [example A]

1. Enter the original pairs in cells B4-C7

Table 2

	B	C
3	**X**	**Y**
4	1	1
5	2	3
6	3	5
7	4	7

2. Create the generating matrix in cells B14-E17 by typing in cells B14-E14 the formulas *=B3^0*, *=B3^1*, *=B3^2*, *=B3^3*, then select cells B14-E14 and pull down the lower right corner by 3 cells.

Table 3

	B	C	D	E
14	1	1	1	1
15	1	2	4	8
16	1	3	9	27
17	1	4	16	64

3. Compute the inverse M^{-1} by typing in cell B26 the formula *=MINVERSE(B14:E17)*, select cells B26-E29, press key F2 followed by Ctrl-Shift-Enter.

Table 4

	B	C	D	E
26	4.000	-6.000	4.000	-1.000
27	-4.333	9.500	-7.000	1.833
28	1.500	-4.000	3.500	-1.000
29	-0.167	0.500	-0.500	0.167

Example: Excel implementation [A] - cont

4. Compute the coefficients C by typing in cell B31 the formula =*MMULT(B26:E29,C4:C7)*, select cells B31-B34, press key F2 followed by Ctrl-Shift-Enter.

Table 5

	B
31	-1
32	2
33	0
34	0

2. From C create the engendering formula:

$$Y = -1 + 2 * X$$

Cells B31-B34 contain the coefficient of X^0, X^1, X^2, and X^3.

3. For any value of X compute the corresponding Y by replacing X in the formula above. For example:

X = 8 implies Y=-1+2*8=15

Examples of patterns

Table 6 below shows several examples of patterns found using the Excel implementation.

Table 6

X	Ya	Yb	Yc	Yd
1	1	1	1	1
2	3	3	3	3
3	5	5	5	5
4	7	7	7	7
5	9	9.96	15.506	716.73
6	11	15.80	48.590	8707.58
7	13	27.40	140.952	58473.76
8	15	48.60	348.976	282983.99

The engendering formulas are:

$$Ya = 2 * X - 1$$

$$Yb = -0.04 + 1.4 * X^2 - 0.4 * X^3 + 0.04 * X^4$$

$$Yc = 0.4458 + 0.8735 * X^3 - 0.3614 * X^4 + 0.042168 * X^5$$

$$Yd = 0.9095 + 0.0945 * X^5 - 3.9638 * 10^{-3} * X^8 + 4.0352 * 10^{-5} * X^{11}$$

Example: R implementation [A]

1. Enter the initial pairs

2. Create the generating matrix

3. Find the constants C

   It produces the same values as in Excel

4. From the model compute Y for X in the range 1-8

   The result are: 1, 3, 5, 7, 9, 11, 13, 15

```
> X <- matrix(c(1,2,3,4),ncol=1)
> Y <- c(1,3,5,7)
>
> M <- cbind(X^0,X^1,X^2,X^3)
>
> C=solve(M,Y)
> C
[1] -1  2  0  0
>
> XI <- matrix((1:8), ncol=1)
> MI <- cbind(XI^0,XI^1,XI^2,XI^3)
> YI <- MI%*%C
```

Example: R implementation (cont.)

Using the R code above we can recreate the same patterns found with Excel, and sometimes experiment with much more complex bases or even arbitrarily chosen functions.

- As a generalization let's choose the next pair as (5, 7) and the base: $X^{0.31}$, $X^{1.54}$, $X^{3.57}$, $X^{3.78}$, $X^{9.11}$. The engendering formula becomes:

$$Y = -0.4356 * X^{0.31} + 1.564 * X^{1.54} - 0.41086 * X^{3.57} + 0.2818 * X^{3.78} - 2.576 * 10^{-6} * X^{9.11}$$

- Another example uses sinusoidal functions:

$$Y = 4.7471 * Sin(X) - 21.6737 * Sin(2 * X) - 43.5023 * Sin(3 * X) - 30.1958 * Sin(4 * X)$$

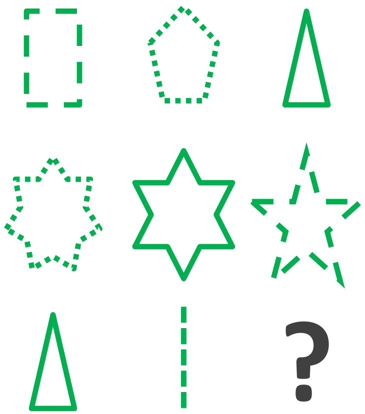
CONCLUSIONS

Mathematical Perspective

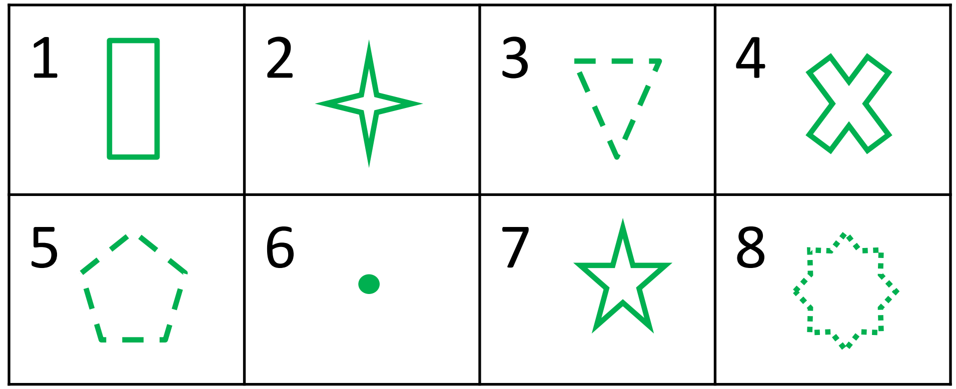
- It is possible and generally easy to find a formula describing the relation between the values of data pairs using simple algebra.
- The discovered pattern is not unique and in most cases it is possible to find a matching formula using almost any type of base and even almost any type of function, not only polynomial.
- It could be easily shown that almost any value could be assigned to the next pair of data.
- The algorithm can be easily expanded for several independent variables.
- If the number of pairs becomes very large this method involves tedious matrix manipulations and the result is too complex to have much practical value.

Qualitative Patterns

Test Set



Options Set



- Most qualitative patterns can be translated into numerical patterns.
- Steps to solve a Test Set: find relevant features, convert features to numbers, find numeric patterns, predict next values, translate numbers into features.
- To avoid the multiple patterns issue tests give an Options Set
- Here the relevant features are number of vertices and line type.

Alternate Reality?

- Since values predicted by each model are so different inside and outside the original range of X it is evident that, at best, a single model is correct although several models can be acceptable approximations.
- This Alternate Reality issue is well known in other fields:
  - Immanuel Kant suggested a division between the natural world called *Noumena* (our X,Y pairs) and the human vision of it called *Phenomena* (model M).
  - Klaus Conrad suggested the term *Apophenia* to describe a "*specific experience of an abnormal meaningfulness*" or our tendency to see patterns in random data.
- To dispel any doubts, the pairs used in this example are totally independent being created by a quasi-random number generator from R (see code snippet)

```
> set.seed(8555)
> C <- floor(t(runif(10))*10)
> Y <- C[1:4]
```

Final Word

- The power and beauty of mathematics could work for or against us depending on how we handle it's intricacies.
- Are you sure you are not gambling the future of your company with the latest analytic fad?
- Contact use today to learn how we can help you validate your analytic processes.

Legal Disclaimer

Under no circumstances but not limited to negligence, shall Innovatorium Technologies Corporation be liable for any direct, indirect, special, incidental or consequential damages whatsoever that result from the use of information presented in this white paper, unless that information is subsequently confirmed in writing as part of a legally binding contract.

The information presented in this white paper cannot and do not address the unique facts and circumstances of your specific situation and should not be relied on for your particular applications. Therefore, you should not use this information without first contacting Innovatorium Technologies Corporation.

Contact Information

Web: www.innovatt.com; www.heuristicanalytics.com

Email: bd @ innovatt.com

Phone: +1-267- 342-2815