DECOMPOSITION OF RELATIONS:

A NEW APPROACH TO CONSTRUCTIVE INDUCTION IN MACHINE LEARNING AND DATA MINING - AN OVERVIEW

Marek Perkowski

Part 1

Logic for Learning

- This is a review presentation that presents work done at Portland State University and associated groups in years 1989 - 2001 in the area of functional decomposition of multi-valued functions and relations.

Contributors

Marek Perkowski

Dr. Alan Mishchenko

Stanislaw Grygiel, Ph.D., Intel
Craig Files, Ph.D., **Agilent.**
Paul Burkey, **Intel**
Rahul Malvi, **Synopsys**
Michael Burns, **Vlsi logic**,
Timothy Brandis, **OrCAD**
Tu Dinh,
Michael Levy, **Georgia Tech**

Prof. Bernd Steinbach, Germany

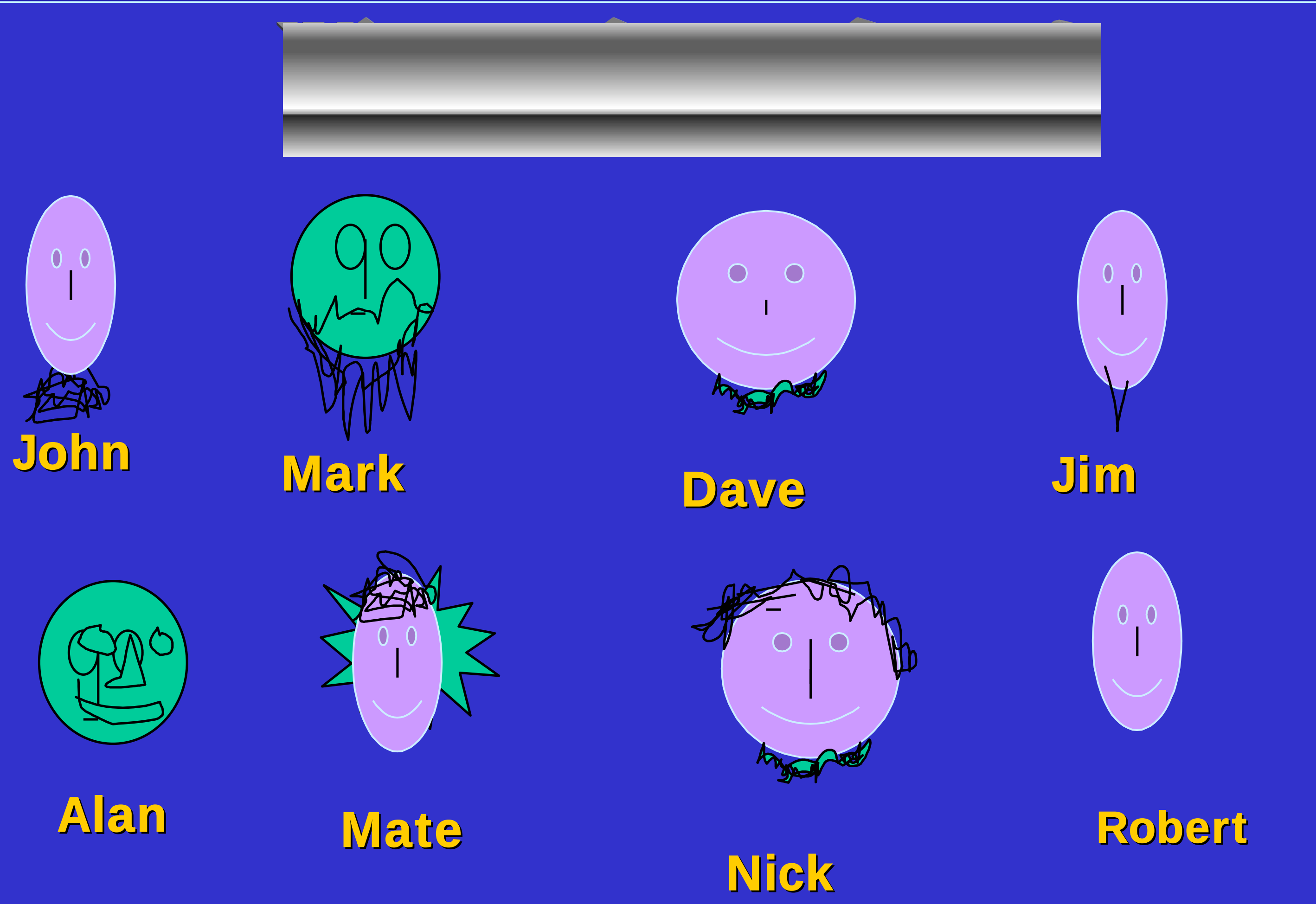
Prof. Lech Jozwiak, The Netherlands

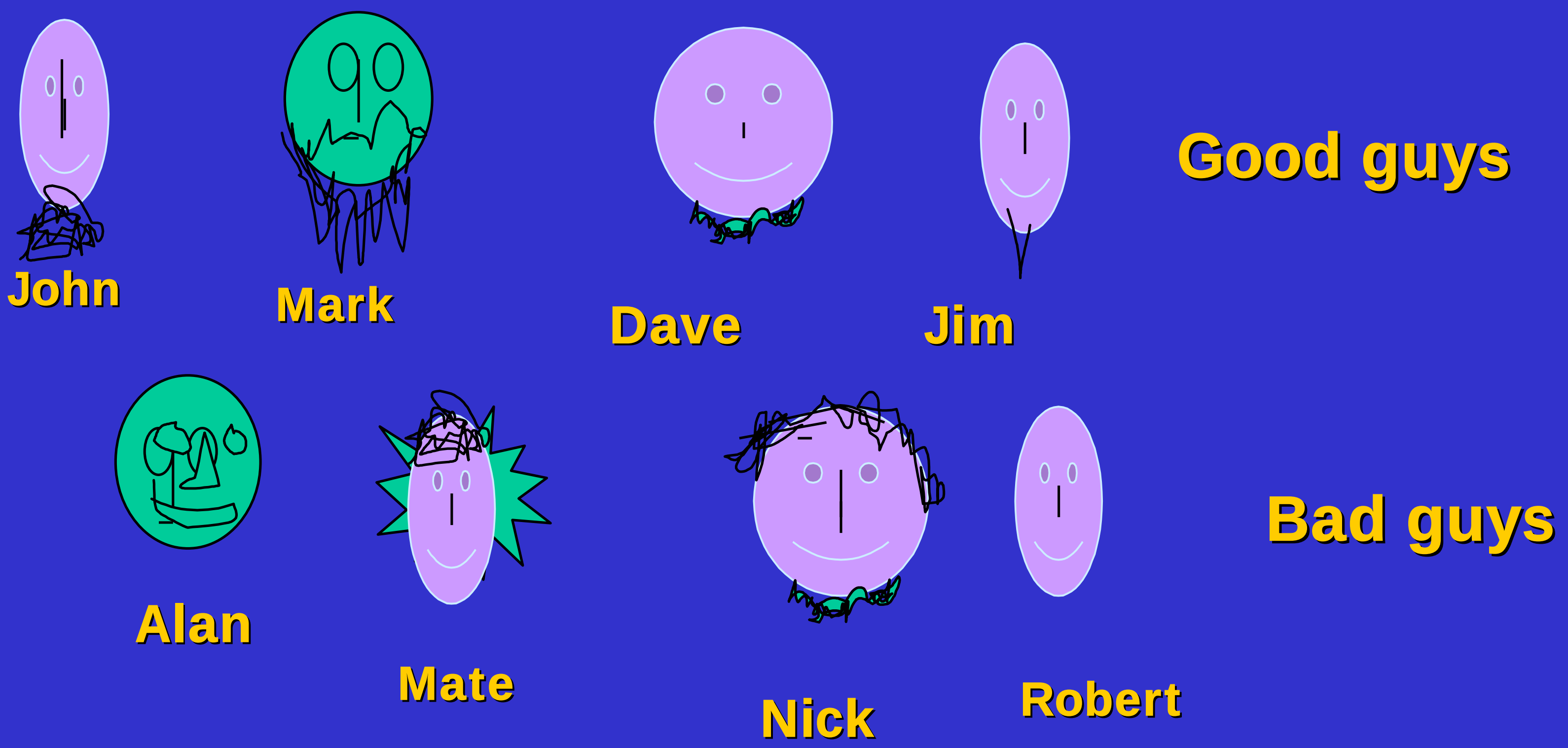
Martin Zwick, System Science PSU

George Lendaris, System Science PSU

Essence of logic synthesis approach to learning

Example of Logical Synthesis



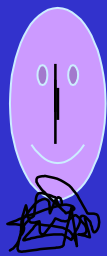


A - size of hair

B - size of nose

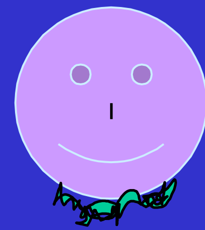
C - size of beard

D - color of eyes

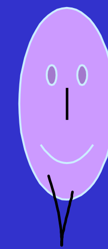


John

Mark



Dave



Jim

Good guys

A’ BCD   A’ BCD’   A’ B’CD   A’ B’CD

AB \ CD	00	01	11	10
00	-	-	1	-
01	—	-	1	1
11	-	—	-	-
10	-	-	-	-

A - size of hair

B - size of nose

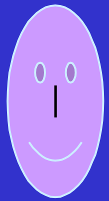
C - size of beard

D - color of eyes

Alan

Mate

Nick



Robert

Bad guys

A' BC'D' AB'C'D ABCD A' B'C'D

AB \ CD	00	01	11	10
00	-	-	1	-
01	0	0	1	1
11	-	-	0	-
10	-	0	-	-

A'C

A - size of hair

B - size of nose

C - size of beard

D - color of eyes

Generalization 1:

Bald guys with beards are good

Generalization 2:

All other guys are no good

AB \ CD	00	01	11	10
00	-	-	1	-
01	0	0	1	1
11	-	-	0	-
10	-	0	-	-

A - size of hair

B - size of nose

C - size of beard

D - color of eyes

A'C

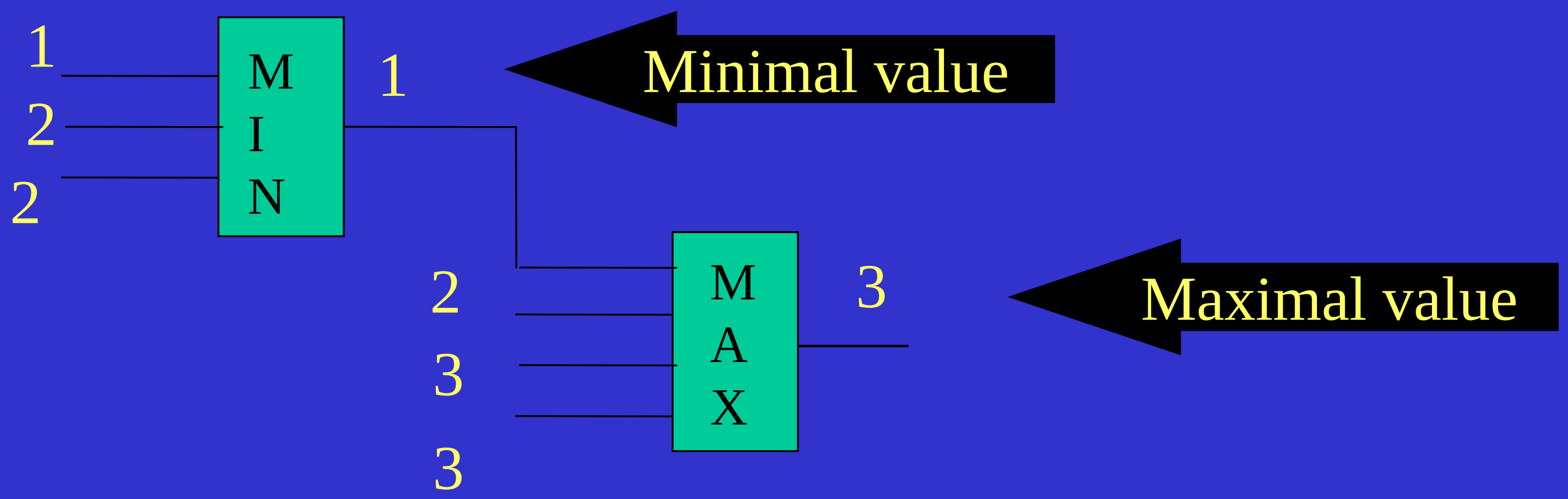
Short Introduction: multiple-valued logic

Signals can have values from some set, for instance {0,1,2}, or {0,1,2,3}

{0,1} - binary logic (a special case)

{0,1,2} - ***a ternary logic***

{0,1,2,3} ***- a quaternary logic,*** **<u>etc</u>**



Types of Logical Synthesis

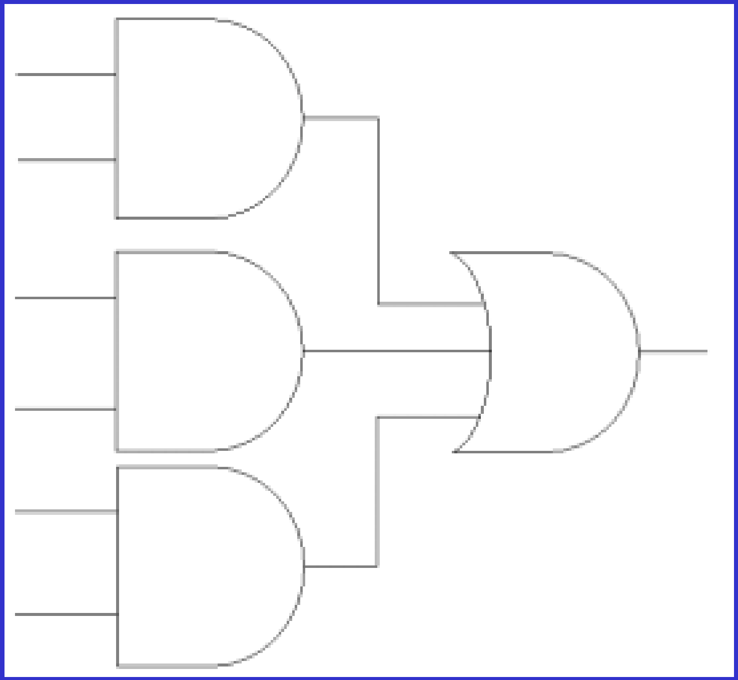
- Sum of Products
- Decision Diagrams
- Functional Decomposition

The method we are using

Sum of Products

Called also SOP or DNF

AND gates, followed by an OR gate that produces the output. (Also, use Inverters as needed.)



Decision Diagrams

A Decision diagram breaks down a Karnaugh map into set of decision trees.

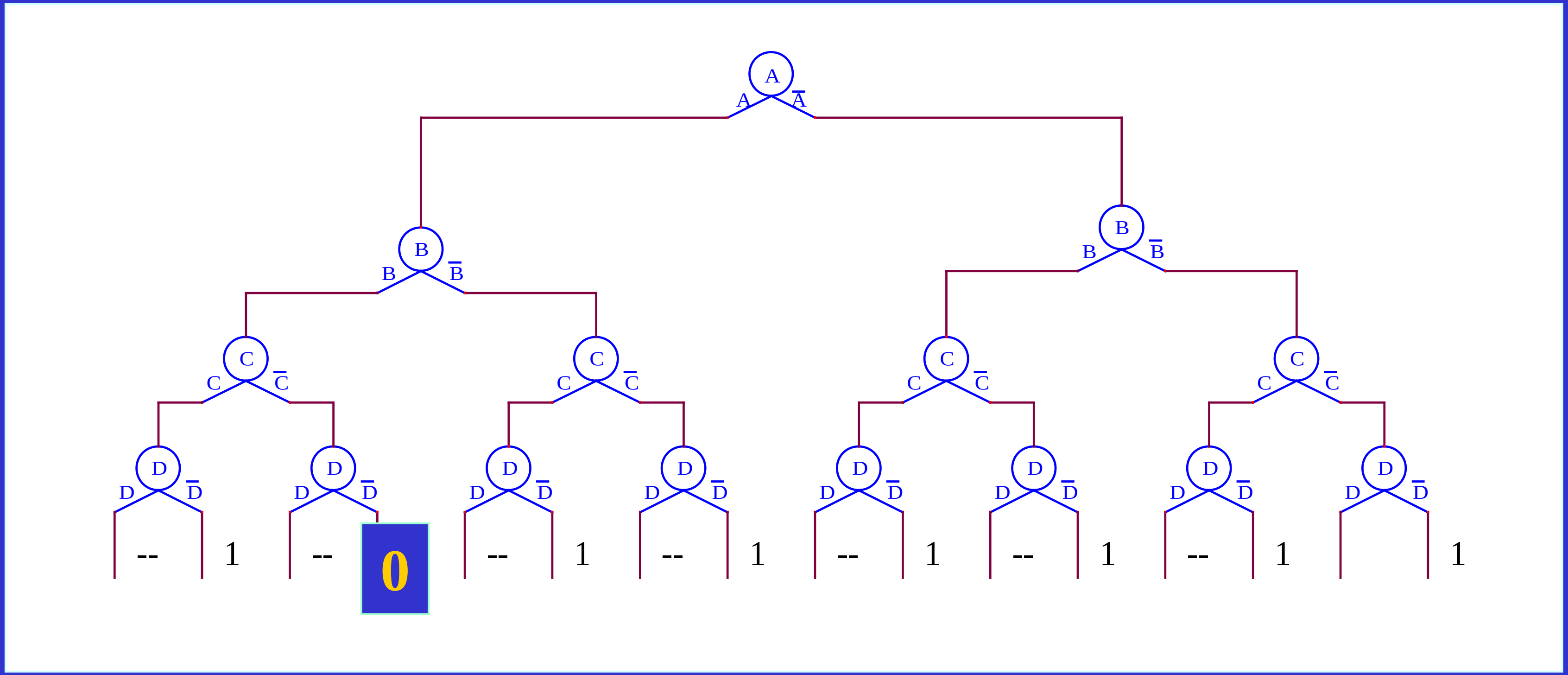
A decision diagram ends when all of branches have a yes, no, or do not care solution.

This diagram can become quite complex if the data is spread out as in the following example.

Example Karnaugh Map

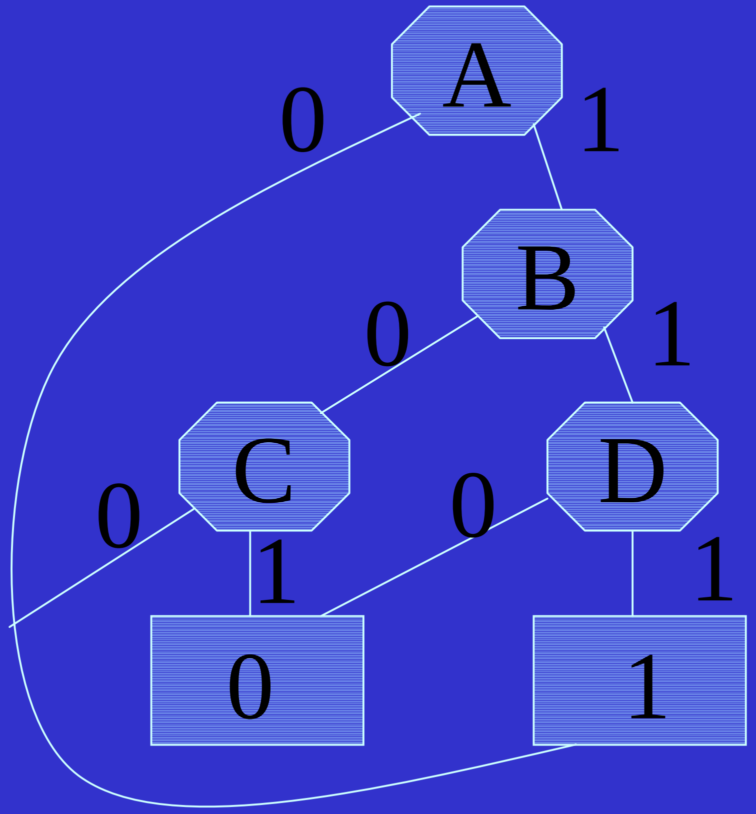
AB\CD	00	01	10	11
00	1	-	1	-
01	-	1	-	1
10	1	-	1	-
11	0	1	-	1

Decision Tree for Example Karnaugh Map



BDD Representation of function

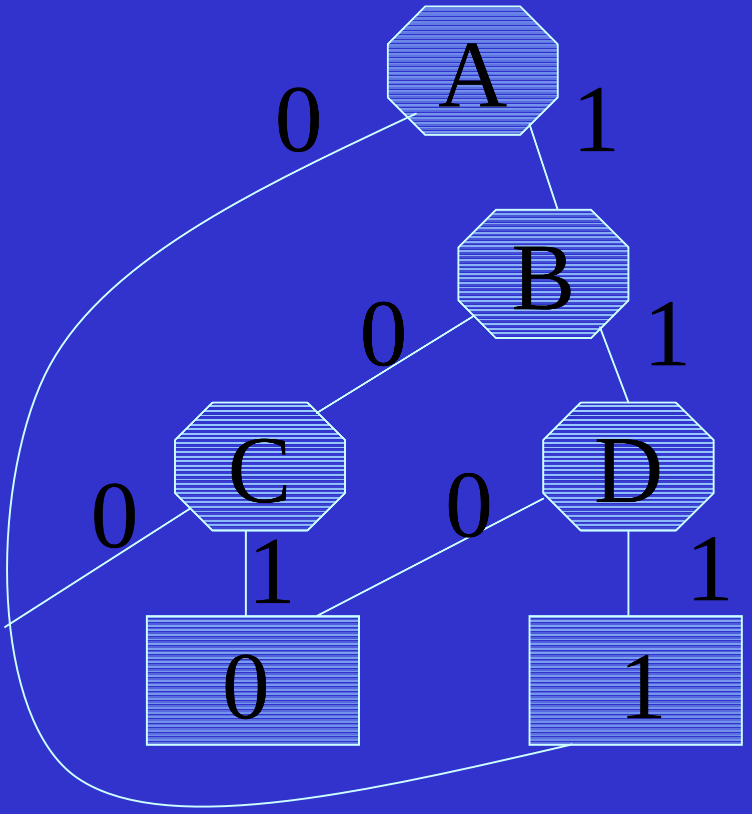
AB \ CD	00	01	11	10
00	-	1	1	-
01	1	-	1	1
11	0	1	-	0
10	-	1	0	-



Incompletely specified function

BDD Representation of function

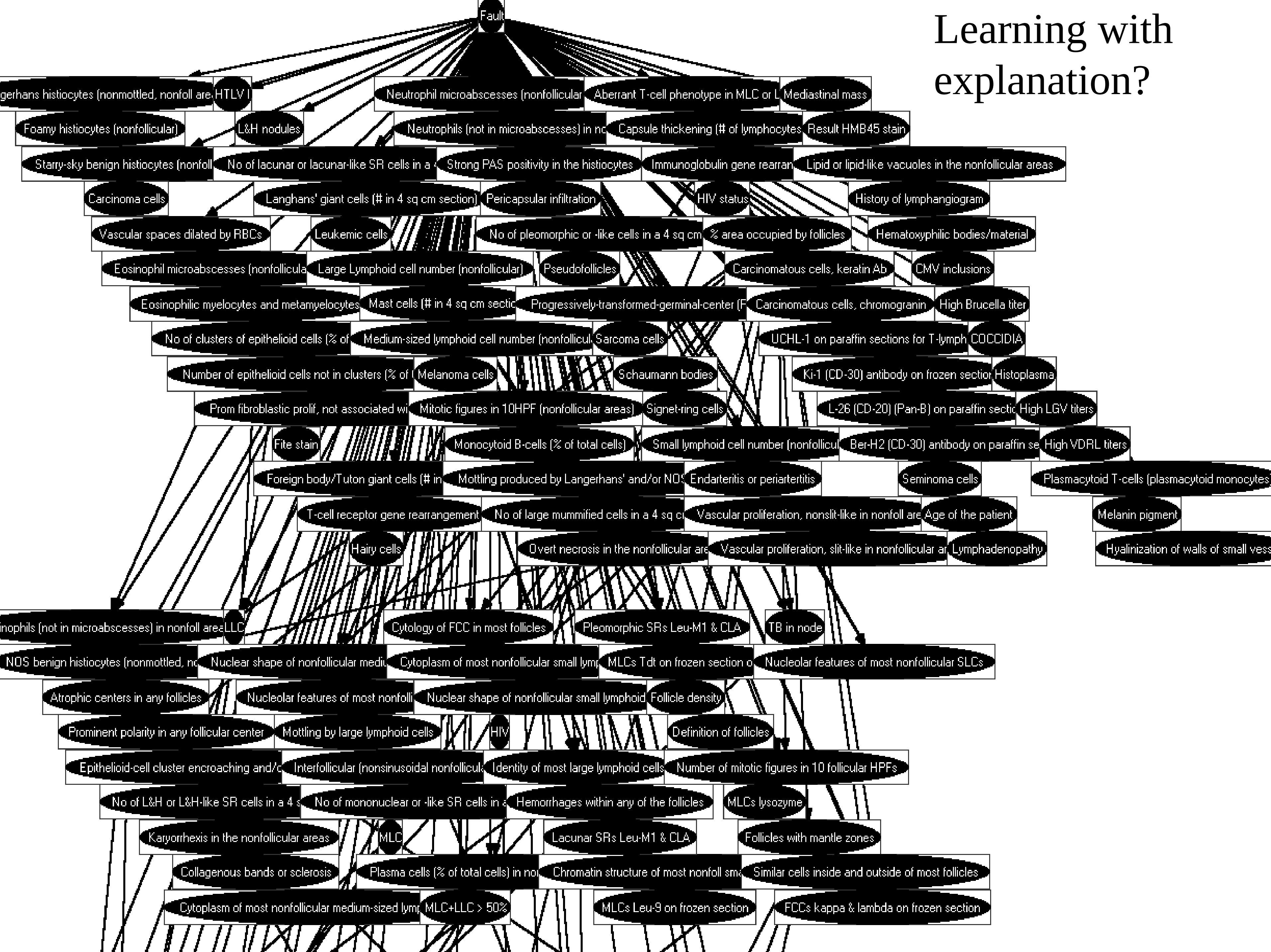
AB \ CD	00	01	11	10
00	1	1	1	1
01	1	1	1	1
11	0	1	1	0
10	1	1	0	0



Completely specified function

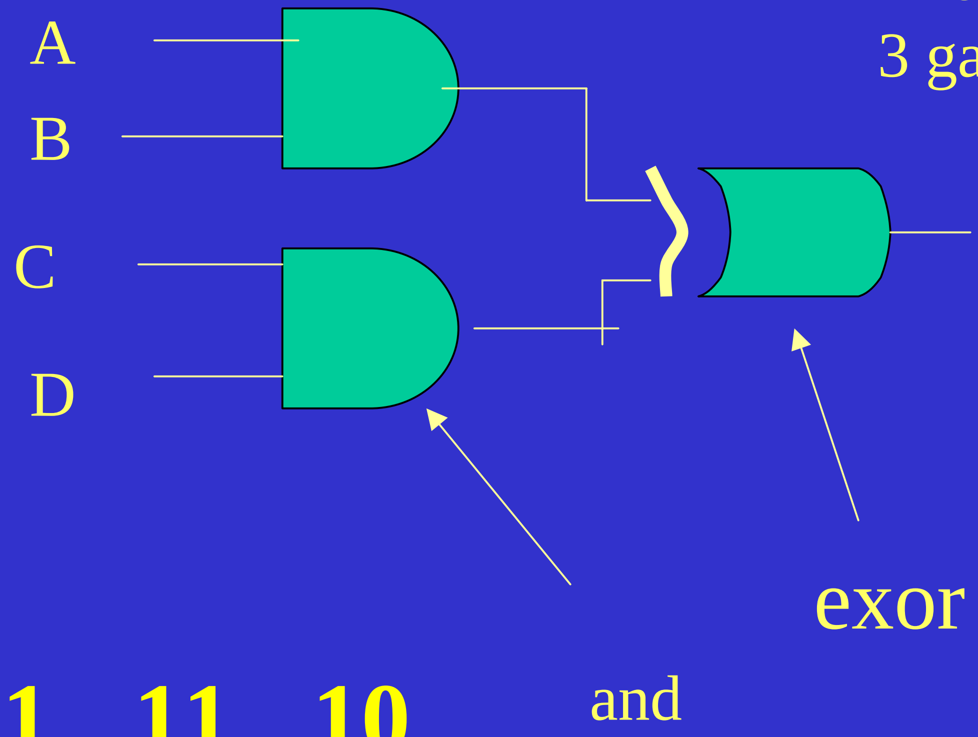
When BDD is finished, it specifies a <u>complete</u> function.

Learning with explanation?



The cost is
3 gates

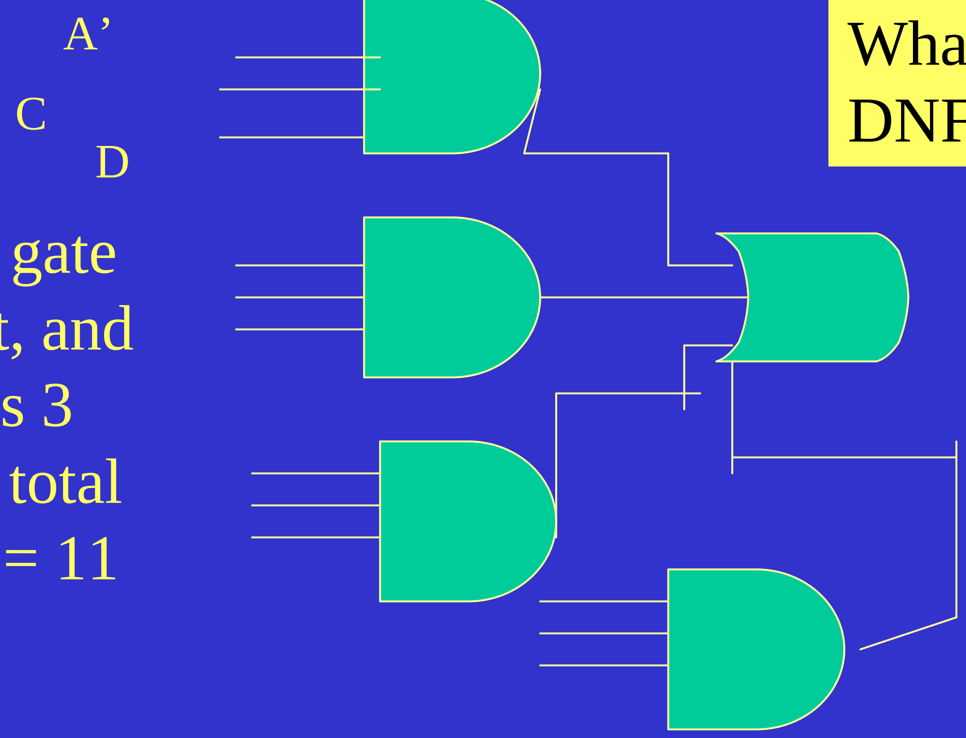
Assuming two-input gates of the same cost, this is the best solution



AB \ CD	00	01	11	10
00	-	-	1	0
01	0	0	1	0
11	1	1	0	1
10	-	0	1	-

Our decomposing program would find this solution

What is wrong with DNF?



Because the 3-input gate needs two to build it, and four input gate needs 3 gates to build it, the total cost is now 4*2 + 3 = 11 instead of 3

AB \ CD	00	01	11	10
00	-	-	1	0
01	0	0	1	0
11	1	1	0	1
10	-	0	1	-

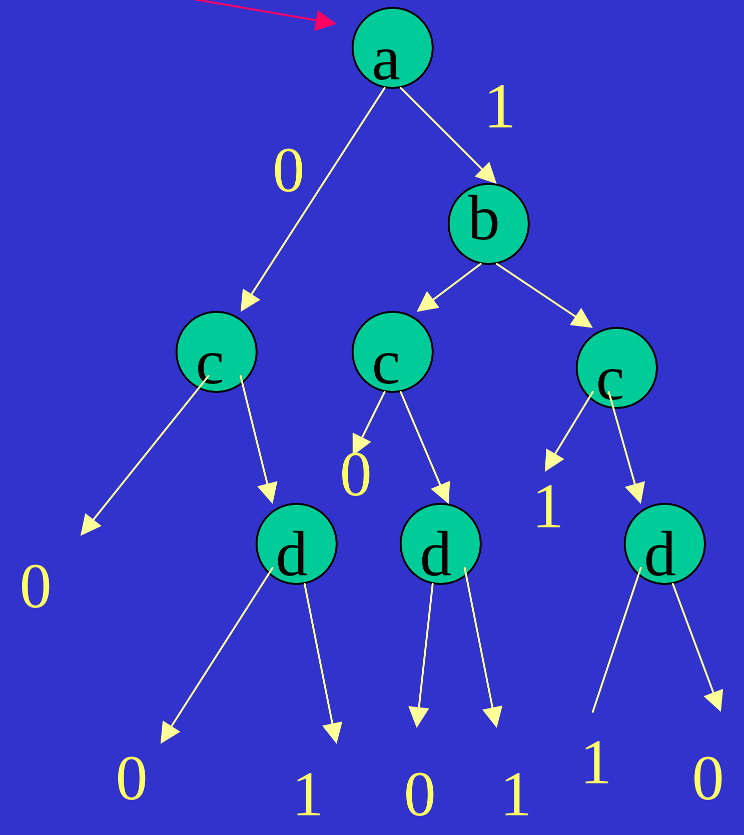
Conclusion: DNF is wasteful, it is used because it is easy and historical

What about Decision Trees and Diagrams that are most often used in industry, data mining and medicine?

This gate costs 3

Total cost is 8 * 3 = 24

AB \ CD	00	01	11	10
00	-	-	1	0
01	0	0	1	0
11	1	1	0	1
10	-	0	1	-

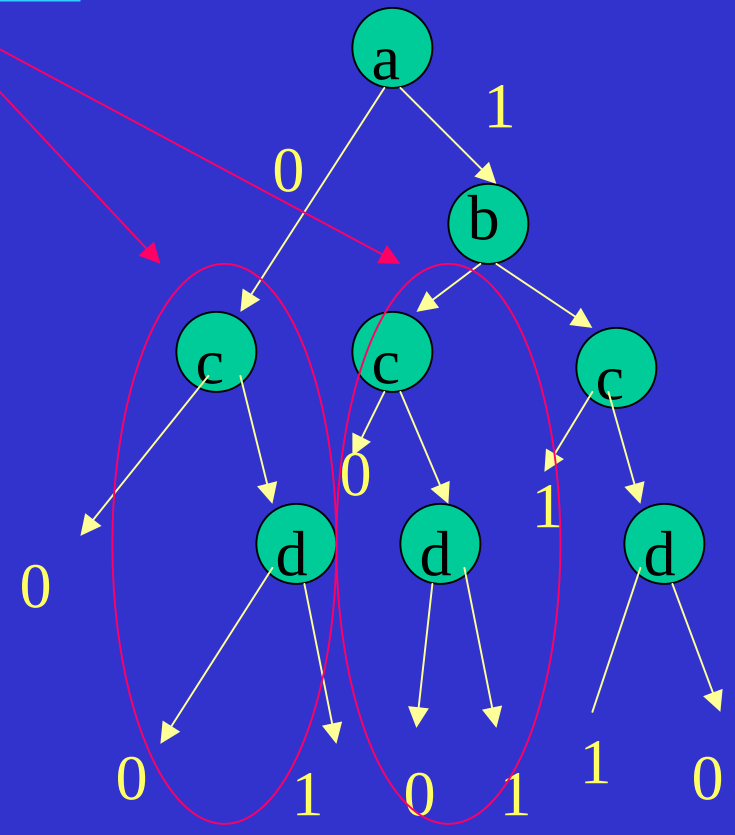


If Decision Trees are no good what about Decision Diagrams?

These subtrees are isomorphic and are combined

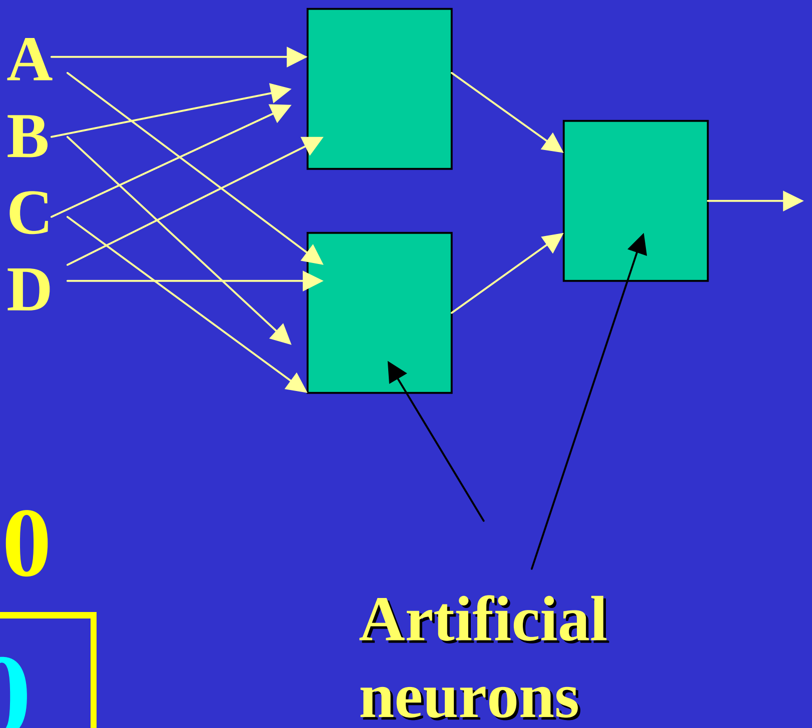
Total cost is 6 * 3 = 18, worse than DNF in this case

AB \ CD	00	01	11	10
00	-	-	1	0
01	0	0	1	0
11	1	1	0	1
10	-	0	1	-



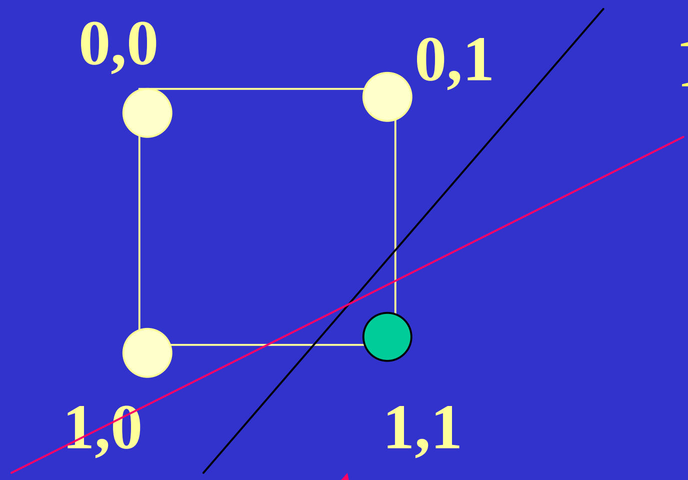
So what about Neural Nets?

There is a very small chance that out of all hypersurfaces separating ones from zeros the one will be found that corresponds to the simplest boolean function



AB \ CD	00	01	11	10
00	_	-	1	0
01	0	0	1	0
11	1	1	0	1
10	-	0	1	-

Why Neural Nets lead to overfitting and it is hard translate the learned network to a boolean function?

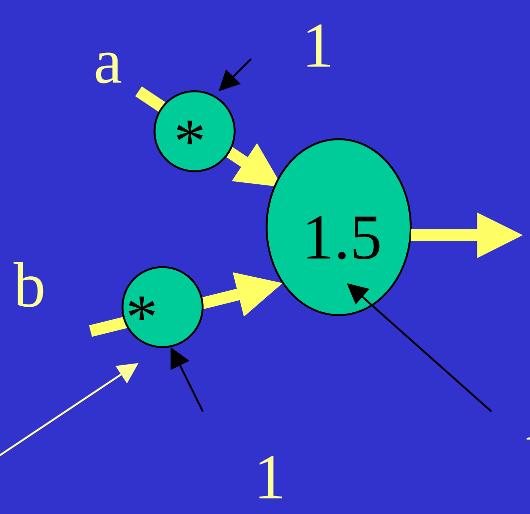


This is function AND

We have practically infinite number of lines separating node (1,1) from other nodes.

For every variable we need a multiplier, we need also an arithmetic adder and comparator.

Such a waste!!



Concluding on current machine learning methods

- DNF, Neural Nets and Decision Trees, Decision Diagrams have their advantages and disadvantages
- Good method should not assume structure of connections or type of gates.
- But will be slower
- Perhaps methods should be combined.

Pros and cons

In generating the final combinational network, BDD decomposition, based on multiplexers, and SOP decomposition, trade flexibility in circuit topology for time efficiency

Generalized functional decomposition sacrifices speed for a higher likelihood of minimizing the complexity of the final network

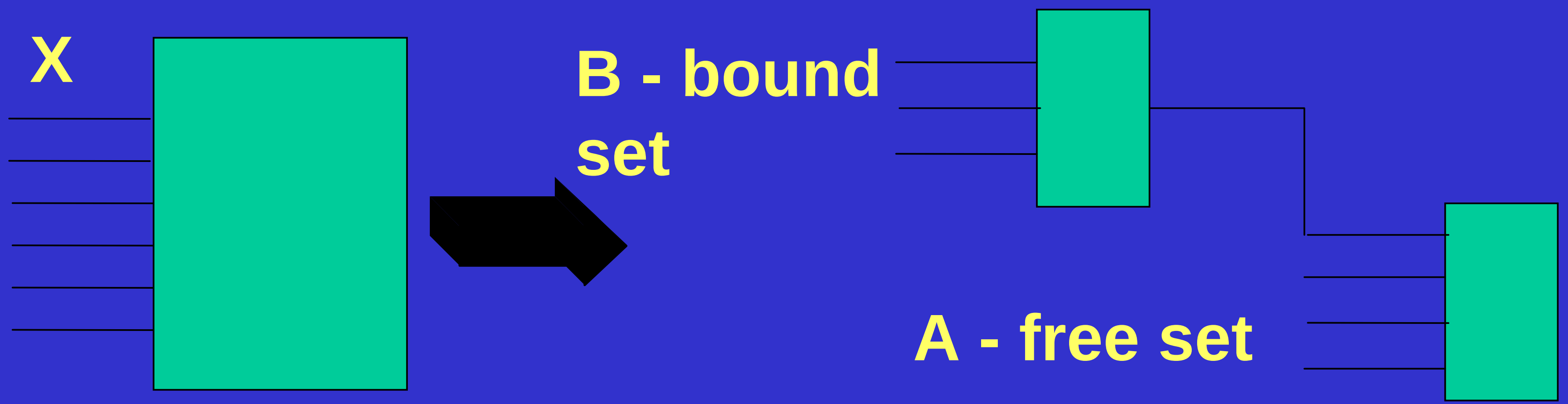
Functional Decomposition

Evaluates the data function and attempts to decompose into simpler functions.

$$F(X) = H(G(B), A),\quad X = A \cup B$$

if $A \cap B = \emptyset$, it is *disjoint decomposition*

if $A \cap B \neq \emptyset$, it is *non-disjoint decomposition*



Applying this decomposition twice, we will get the AB EXOR CD circuit shown earlier.

But this would work only selecting good free and bound sets twice.

Part 2

Overview of data mining

What is Data Mining?

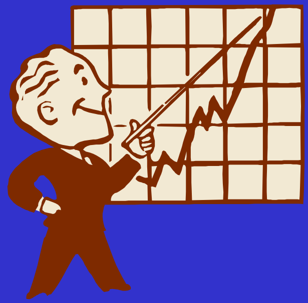
Databases with millions of records and thousands of fields are now common in business, medicine, engineering, and the sciences.

To extract useful information from such data sets is an important practical problem.

Data Mining is the study of methods to find useful information from the database and use data to make predictions about the people or events the data was developed from.

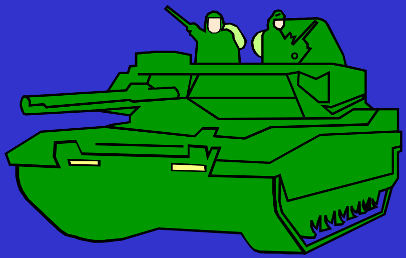
Some Examples of Data Mining

1) Stock Market Predictions



2) Large companies tracking sales

3) Military and intelligence applications



Data Mining in Epidemiology

Epidemiologists track the spread of infectious disease and try to determines the diseases original source

Often times Epidemiologist only have an initial suspicions about what is causing an illness. They interview people to find out what those people that got sick have in common.

Currently they have to sort through this data by hand to try and determine the initial source of the disease.

A data mining application would speed up this process and allow them to quickly track the source of an infectious diseases

Types of Data Mining

Data Mining applications use, among others, three methods to process data

1) Neural Nets

2) Statistical Analysis

3) Logical Synthesis

The method we are using