1 Input

An $m * (n+1)$ sized data set with

- m many row vectors $a, b, c, d, ...m$
- n many dimensional attributes: $A_1, A_2, ...A_n$
- real valued entries for each dimension $1...n$ in each vector
- a binary value $0, 1$ assigned to each vector

2 Procedure

1. Classify vectors into two sets $S+$ and $S-$ based on their binary value of $0, 1$ so that everything of value 0 is in $S+$ and 1 in $S-$.
2. For each pair (a, b), $a \in S+$, $b \in S-$, we compute the difference vector $\delta_{a,b} = (|a_1 - b_1|, |a_2 - b_2|, ...)$. Let the set of these difference vectors be Δ.
3. We use Δ to find the set of vectors Min_Δ such that for each vector $min_\delta \in Min_\Delta$ there is no vector $\delta \in \Delta$ such that each coordinate of δ is less than or equal to each coordinate of min_δ.
4. Min_Δ is the input for a Dualization process which will output a set ε of vectors ϵ that are the combinations of out highest tolerance of error in our attributes.
5. We sort ε by variance methods and choose those ϵ which look most promising for a simple end formula.
6. For such a promising ϵ we compute a table showing the columns and attributes that are needed to distinguish $S+$ from $S-$ withstanding an error of ϵ.
7. Using this small table, we try to find the simplest functions that distinguish $S+$ from $S-$.

3 Output

- A combination of maximal $\epsilon\ 's$ and relevant attributes that produce useful formulas for diagnosing a new patient as $S+$ or $S-$.