Australian National University

# Hierarchical Fuzzy Systems

COMP4660/8420 - Neural Networks, Deep Learning and Bio-inspired Computing

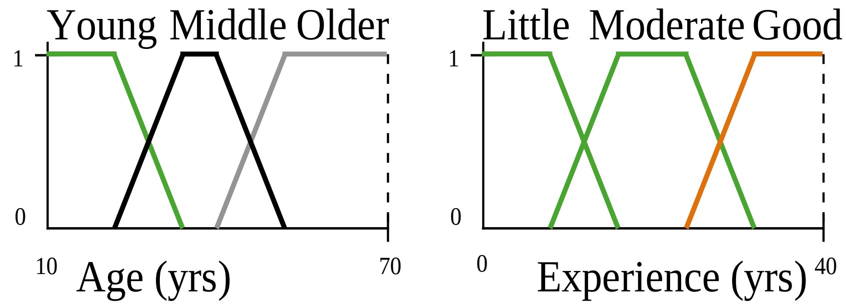
Human Centred Computing

# Overview

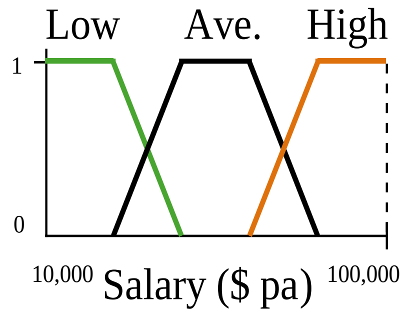
- How do fuzzy systems work?
  - Dense fuzzy rule bases
- Problem definition
  - $|R| = O(T^k)$
- Sketch of solution
- Sparse rules - fuzzy interpolation
  - Interpolation overview
  - Conservation of fuzziness
- Hierarchical dense rule bases
  - Input contributions
- Hierarchical (sparse) rule bases
  - Case study

# Dense fuzzy rule bases

IN:



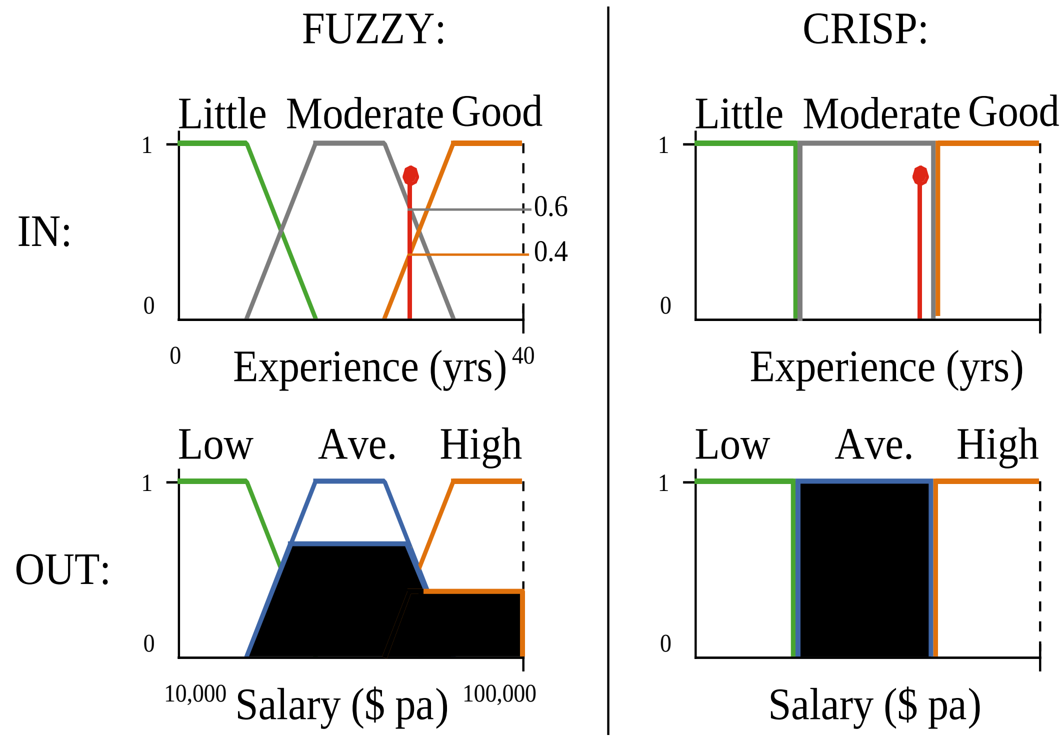

OUT:




- Rules: age & experience to salary
  - IF Age=Young & Exp=Little THEN $=Low
  - IF Age=Young & Exp=Moderate THEN $=Low
  - IF Age=Young & Exp=Good THEN $=Ave
  - …
  - IF Age=Older & Exp=Moderate THEN $=Ave
  - IF Age=Older & Exp=Good THEN $=High
- Three terms, two inputs ⇒ 9 rules

# Fuzzy reasoning

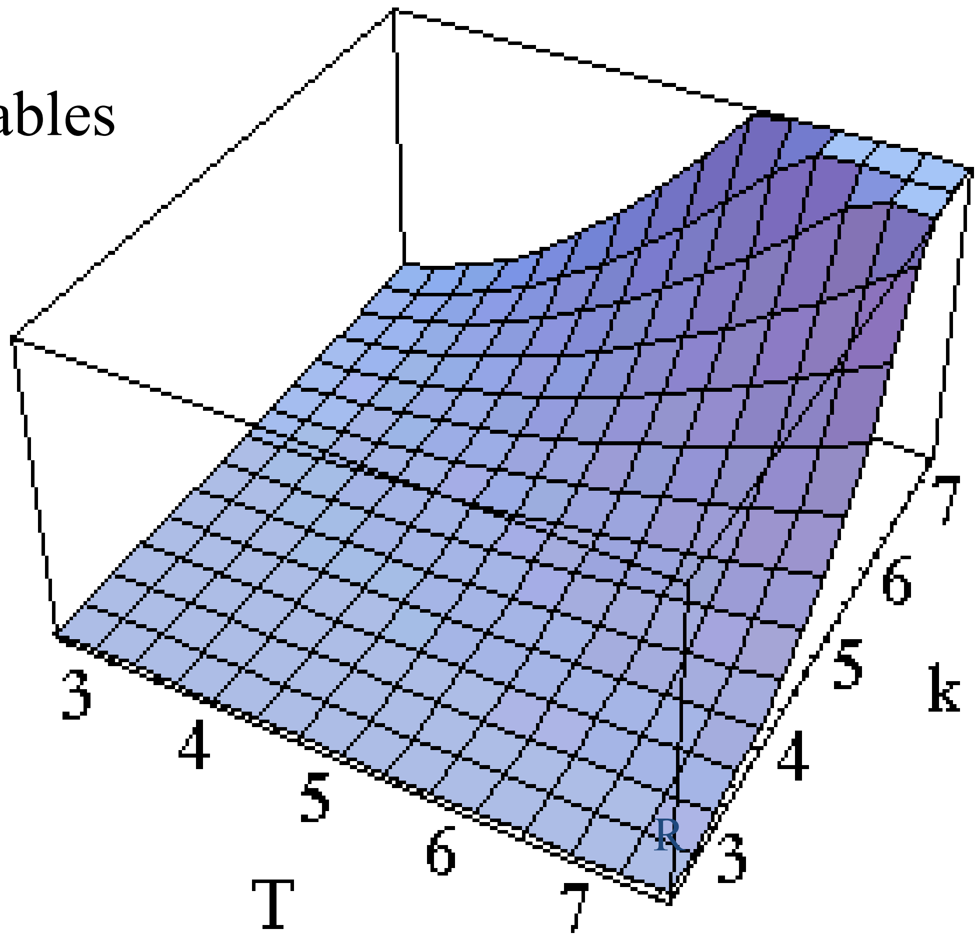
- For Age=Middle, & Exp near border:
  - IF Age=Middle & Exp=Mod. THEN $=Ave
  - IF Age=Middle & Exp=Good THEN $=High




- Fuzzy value experience is Moderate = 0.6, Good = 0.4
  - Result will share properties of both Ave. and High salary range
- Crisp version not very good!

# Problem Definition

- Problem – number of rules:
  - We use T terms
  - For each of k input variables
- $|R| = O(T^k)$
  - To solve real problems $\Rightarrow$ many rules required!
  - E.g. 5 terms, 5 inputs $\Rightarrow$ 3,125 rules

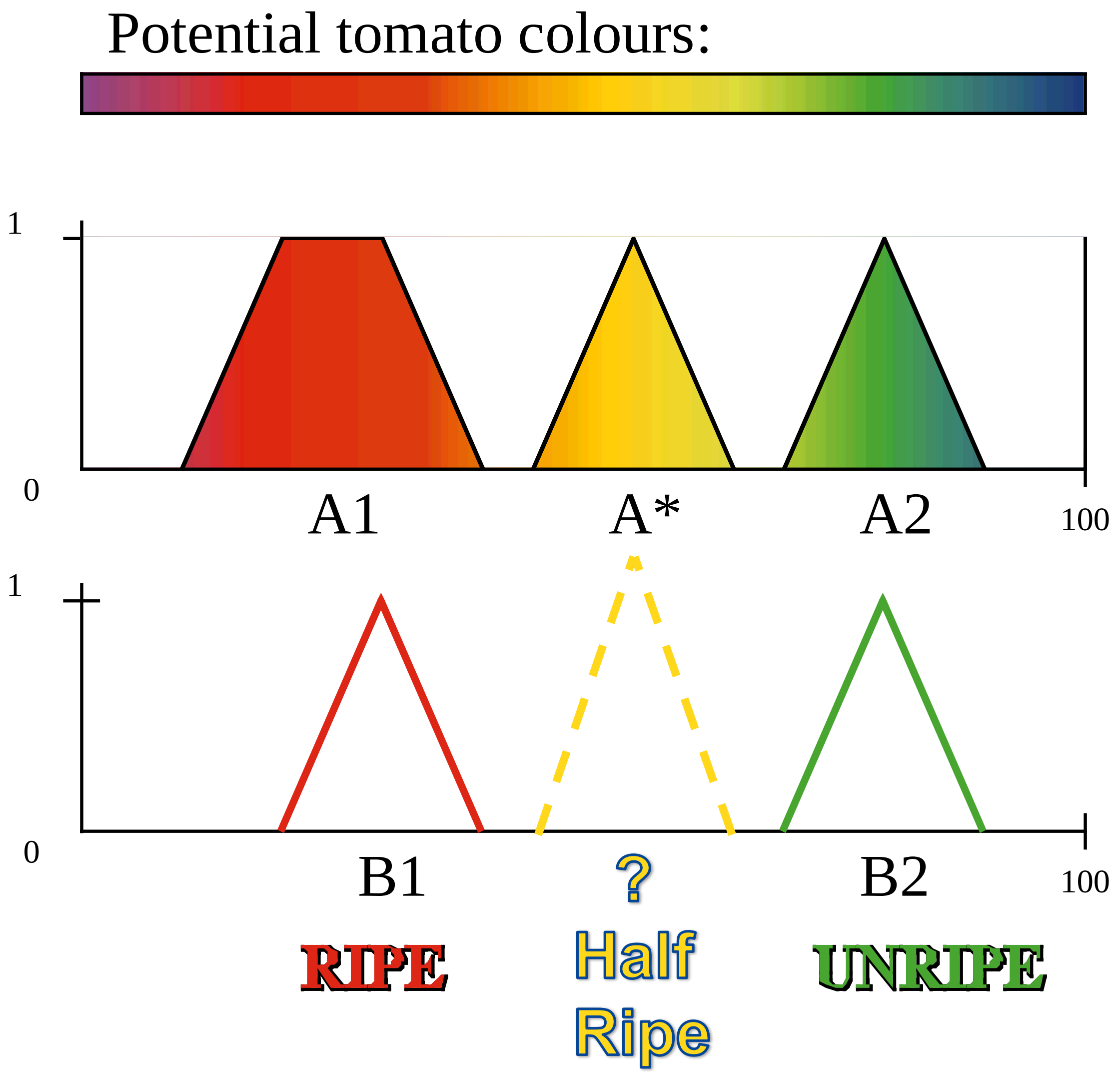


# Sketch of Solution

- Only 3 possible solutions:
  - decrease T,
  - decrease k, or
  - decrease both.
- Decrease T
  - allow sparse fuzzy rule bases
  - require reasoning technique for cases where no rule holds
    - use nearby rules - fuzzy interpolation
- Decrease k:
  - hierarchical fuzzy rule bases
    - often full cover in bordering domains so complexity is not reduced
- Decrease T and k:
  - hierarchical sparse fuzzy rule bases
    - interpolate between different branches of hierarchical rule tree
    - can omit bordering domains and just interpolate from nearby rules

# Interpolation overview

- Tomato colours:
  - IF colour = Red THEN its Ripe
  - IF colour = Green THEN its Unripe
- What about a yellow tomato?
- This is an obvious solution now!

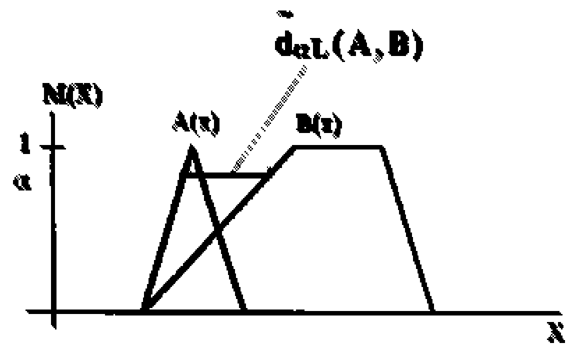


# Fuzzy Distance

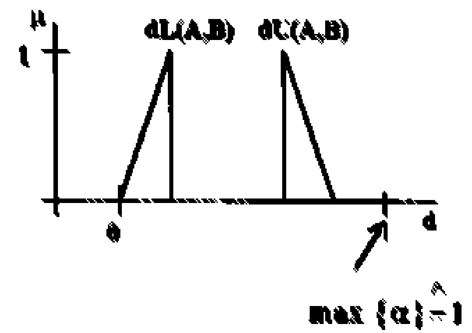
- Fuzzy distance of comparable fuzzy sets:

  for all α∈[0,1] is the pairwise distances between the two extrema of these fuzzy sets ("lower" and "upper fuzzy distance" of the two α-cuts)

DISTANCE OF COMPARABLE FUZZY SETS: A≺B




LOWER AND UPPER DISTANCE
OR
CENTRAL (MEAN) DISTANCE AND
WIDHT (FOR BOTH SETS)




THE FUZZY DISTANCE IS A FUZZY SET OF DISTANCES.

# Fundamental equation of linear interpolation and its solution for B*

- We assume $A_1 \prec A^* \prec A_2$ and $B_1 \prec B_2$
- Distances:

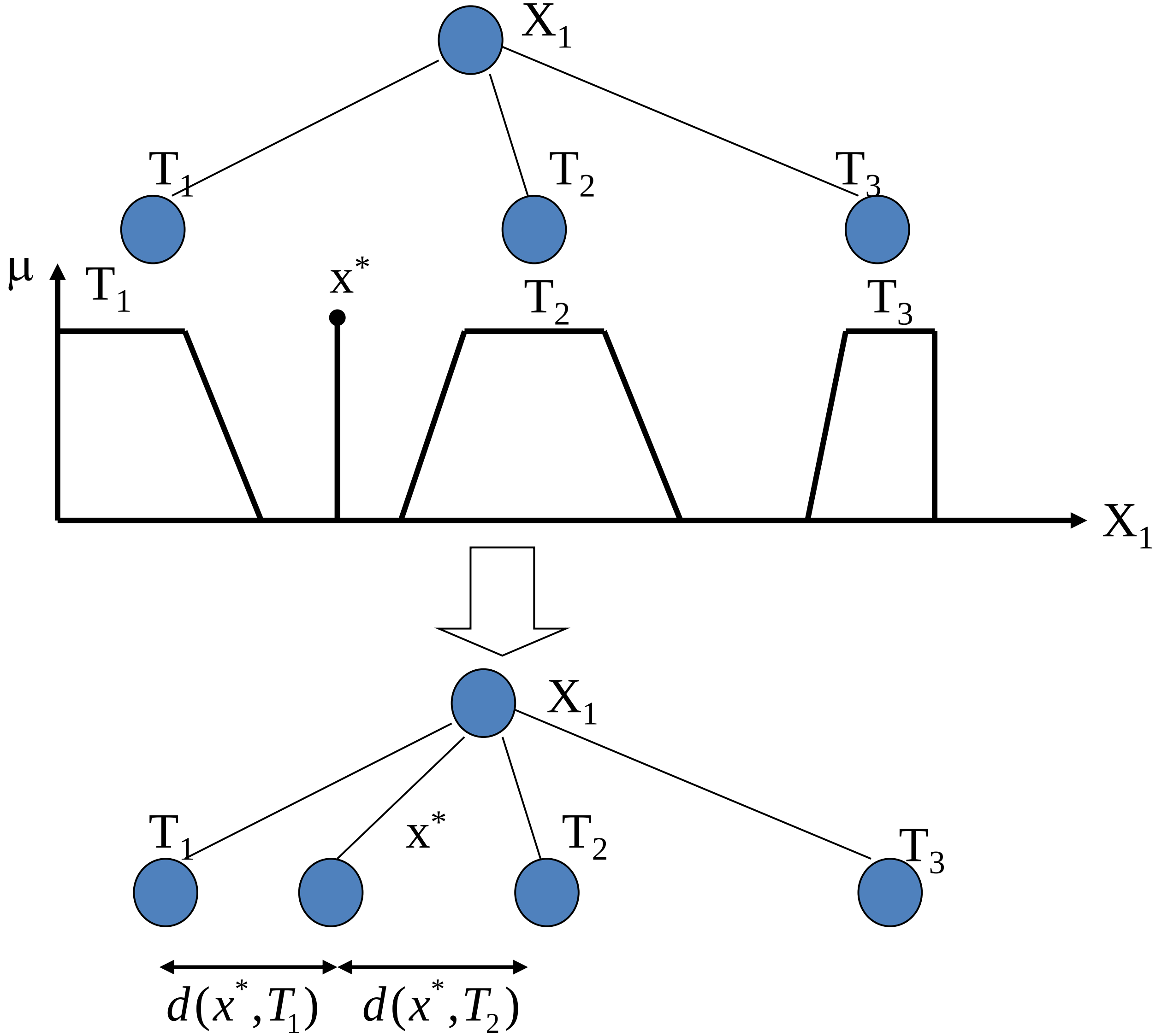
$$D(A^*, A_1) : D(A^*, A_2) = D(B^*, B_1) : D(B^*, B_2)$$

where:

$$\inf\{B_\alpha^*\} = \frac{\dfrac{\inf\{B_{1\alpha}\}}{d_{\alpha L}(A_{1\alpha}, A_\alpha^*)} + \dfrac{\inf\{B_{2\alpha}\}}{d_{\alpha L}(A_{2\alpha}, A_\alpha^*)}}{\dfrac{1}{d_{\alpha L}(A_{1\alpha}, A_\alpha^*)} + \dfrac{1}{d_{\alpha L}(A_{2\alpha}, A_\alpha^*)}}$$

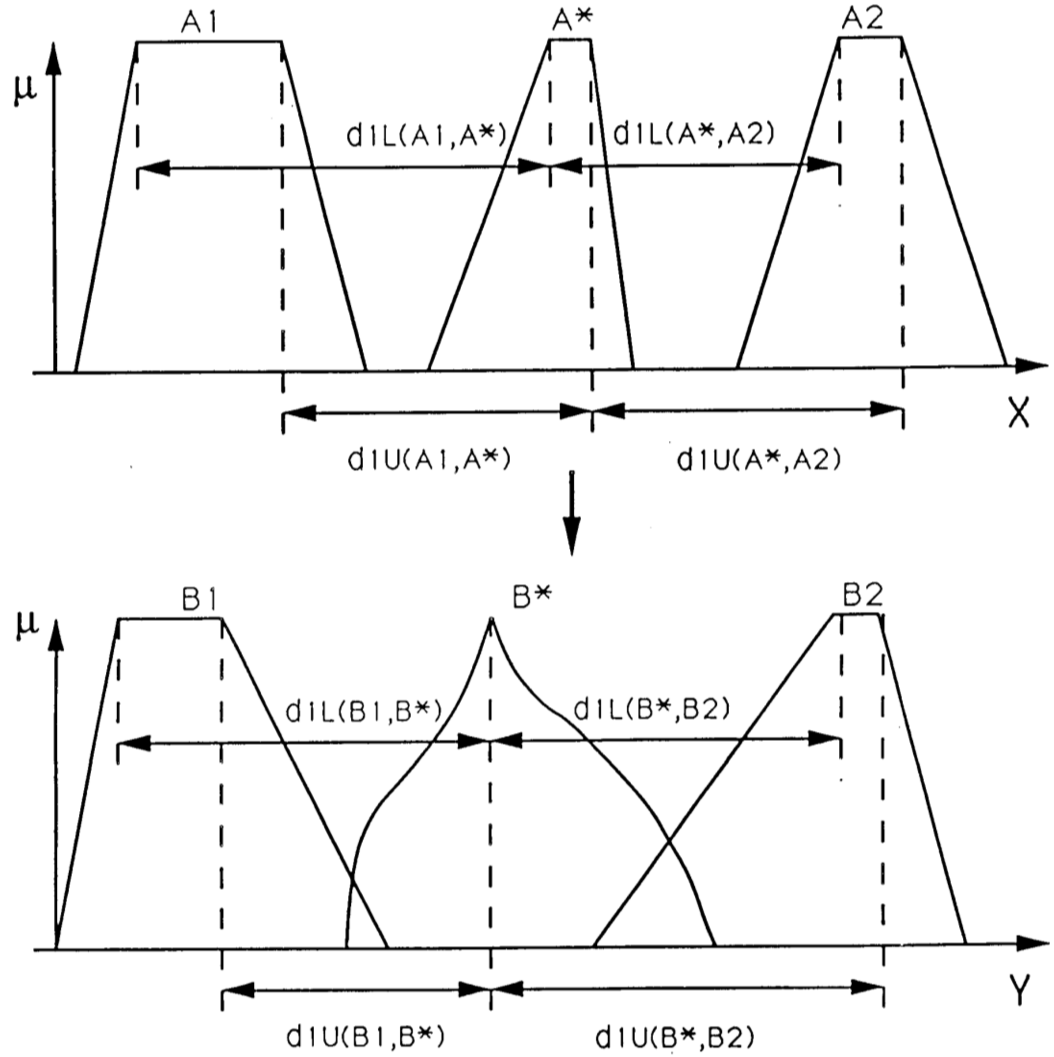
$$\sup\{B_\alpha^*\} = \frac{\dfrac{\sup\{B_{1\alpha}\}}{d_{\alpha U}(A_{1\alpha}, A_\alpha^*)} + \dfrac{\sup\{B_{2\alpha}\}}{d_{\alpha U}(A_{2\alpha}, A_\alpha^*)}}{\dfrac{1}{d_{\alpha U}(A_{1\alpha}, A_\alpha^*)} + \dfrac{1}{d_{\alpha U}(A_{2\alpha}, A_\alpha^*)}}$$

# Example



# Result for the linear interpolation method

- Exact method is expensive to calculate and expensive to use
- Generally just use the 'core' points
  - the four points which define the trapezoid or triangle (2 pts are same for triangle)



# Fuzzy interpolation

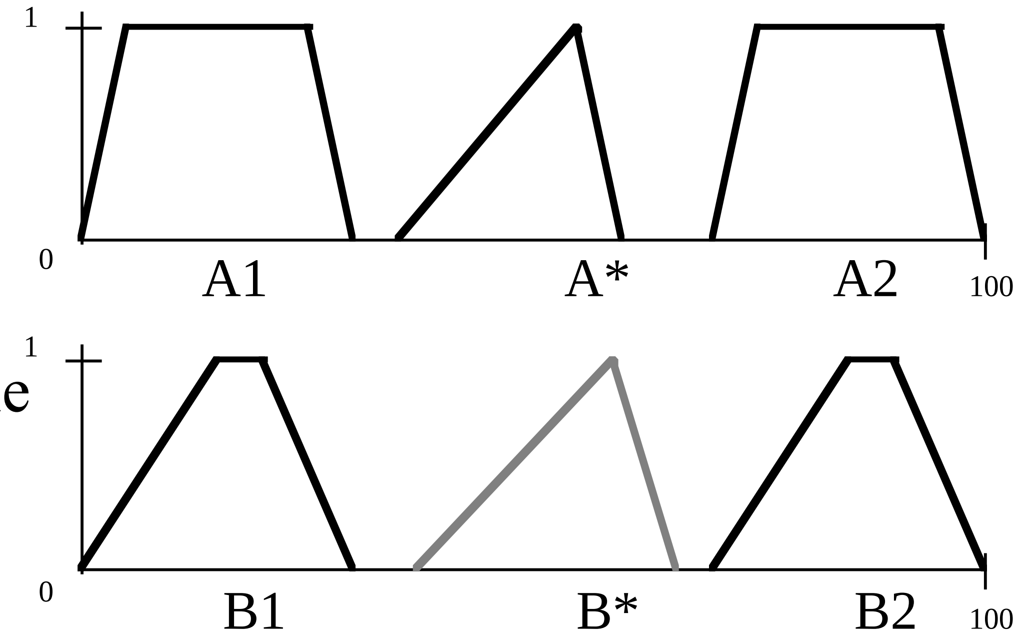
- Sparse rule bases because ...
  - Information is not available
  - Availability cost or natural gaps
  - Deliberate reduction for efficiency
- All methods are descendants of Kóczy & Hirota (1990, 1993) linear interpolation
  - Reduced computational cost
  - But can lead to distorted / abnormal fuzzy rules
- Conservation of fuzziness
  - Only near points of rules used
  - Fuzziness can only increase
  - Core of B* by linear interpolation of near core points

# Interpolation overview 2

- Fuzzy rule based systems
  - used in applications where approximate reasoning is required
  - sparse rule bases
    - information is not available, availability costs or natural gaps
- Fuzzy rule interpolation
  - provide conclusions where
    - no overlap with even the supports of existing rules in the rule base
    - descendants of Kóczy & Hirota (1990, 1993) linear interpolation
  - advantages / disadvantages
    - reduced computation cost / can lead to abnormal fuzzy rules
  - conservation of fuzziness method
    - always acceptably formed rules
    - (additively) conservative – use degree of local fuzziness
    - use the nearby sides of rules, no handedness of rules

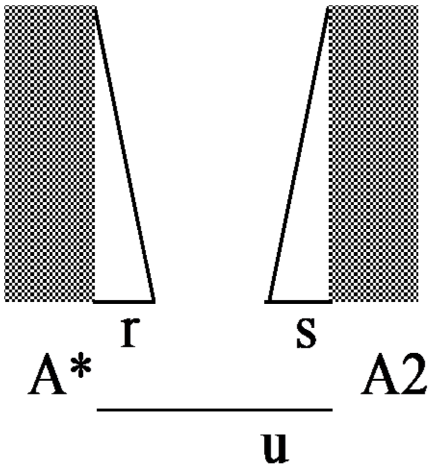
# Conservation of fuzziness

- Assume little homogeneity in the rule base:
  - only nearest core points are visible
  - i.e. A* is in a valley between A1 and A2
  - core of B* from simple linear interpolation between the nearest core points of A1, A2 and B1, B2

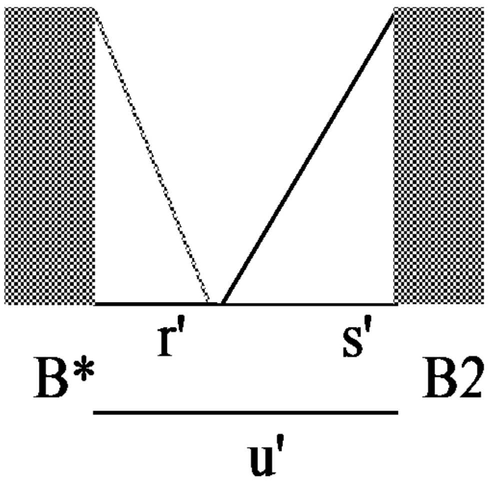



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A1 : | 0, | 5, | 25, | 30 | A* : | 35, | 55, | 55, | 60 |
| A2 : | 70, | 75, | 95, | 100 | B* : | 37, | 59, | 59, | 66 |
| B1 : | 0, | 15, | 20, | 30 | | | | | |
| B2 : | 70, | 85, | 90, | 100 | | | | | |

# Conservation of fuzziness 2




- Spread represents fuzziness of
  - s, s' - rule antecedent, consequent
  - r, r' – observation, conclusion
  - u, u' - A*,A2 distance, B*,B2 distance
- Intuition
  - B2 is more fuzzy than A2.
- Calculating r'
  - Increase in relative local fuzziness




$$r' = r \cdot \frac{u'}{u} \cdot \left(1 + \frac{s' - s''}{z}\right)$$

(where z is s' or s")

$$s'' = s \cdot \frac{u'}{u}$$

# Conservation of Fuzziness Additive strategy

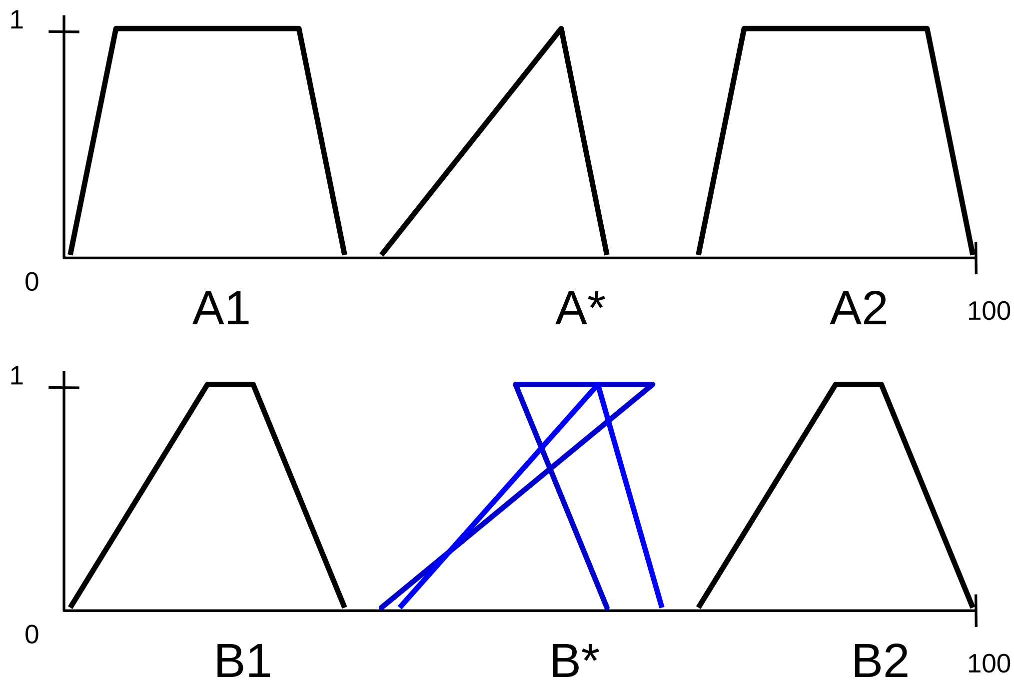
- Restrict notion of local fuzziness
  - only increase only from rule antecedent to consequent
  - i.e., where consequent is less fuzzy (steeper slope) than the antecedent, this is not propagated
    - otherwise would imply that knowledge in (sparse) rule base was sufficient to take a highly fuzzy observation and return a less fuzzy conclusion
    - this would be counter-intuitive
- Additive calc. for r'

$$r' = r \cdot \left(1 + \mathrm{pos}\left(\frac{s'}{u'} - \frac{s}{u}\right)\right)$$

  - r' is not dependent on the ratio of the different metrics
    - crisp s or s' no longer a problem
  - s, s' normalised with respect to u, u'

# Results and comparisons

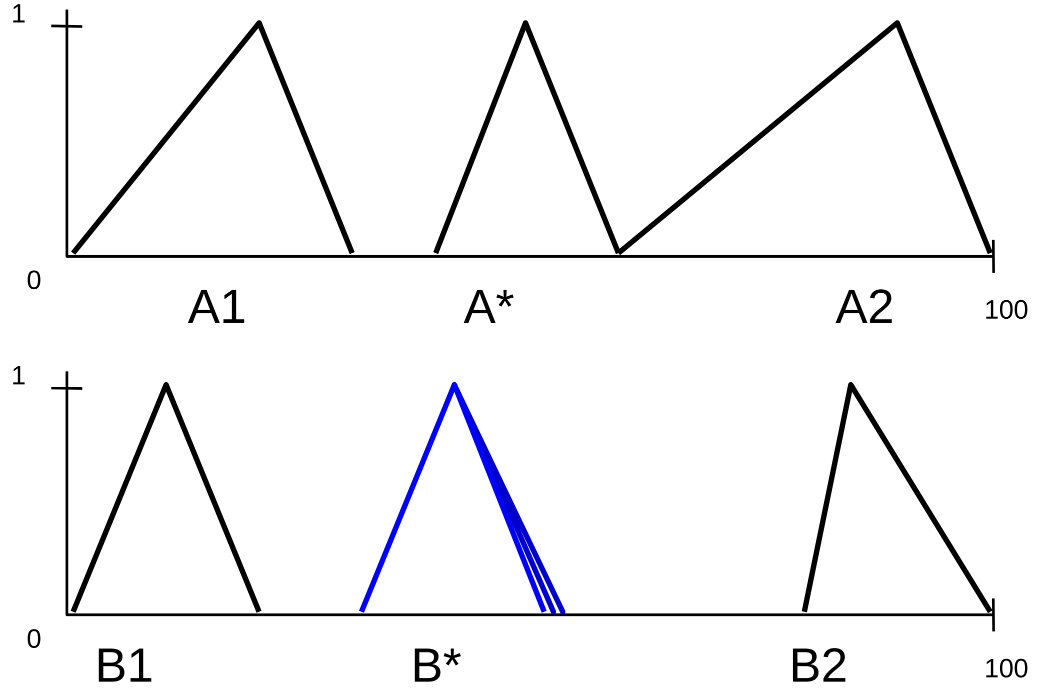
- Example p2
- k & h
  - abnormal conclusion
- g+
  - well formed conclusion
  - note similarity of the left flank results with k & h results (37 versus 35)




| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 : | 0, | 5, | 25, | 30 | A* : | 35, | 55, | 55, | 60 | |
| A2 : | 70, | 75, | 95, | 100 | B* : | 35, | 65, | 50, | 60 | k&h |
| B1 : | 0, | 15, | 20, | 30 | B* : | 37, | 59, | 59, | 66 | g+ |
| B2 : | 70, | 85, | 90, | 100 | | | | | | |

# Results & comparisons – 2

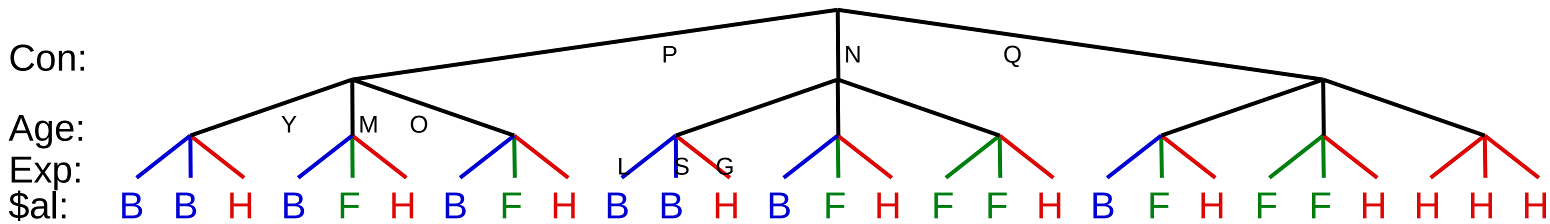
- Example p5
- k & h
  - abnormal conclusion
- g+
  - well formed conclusion
  - note similarity of right flank results with k & h results (52 versus 54)



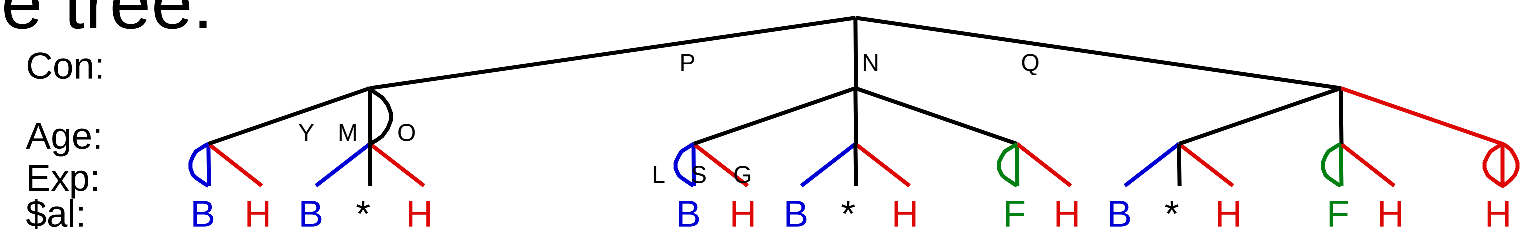

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A1 : | 0, | 20, | 20, | 30 | A* : | 40, | 50, | 50, | 60 | |
| A2 : | 60, | 90, | 90, | 100 | B* : | 53, | 42, | 42, | 54 | k&h |
| B1 : | 0, | 10, | 10, | 20 | B* : | 32, | 42, | 42, | 52 | g+ |
| B2 : | 80, | 85, | 85, | 100 | | | | | | |

# Hierarchical dense rule bases – salary dataset

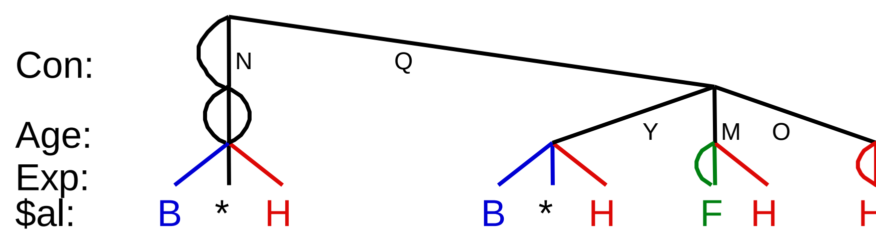
- Rules in a tree (Con/Age/Exp)




- Prune tree:

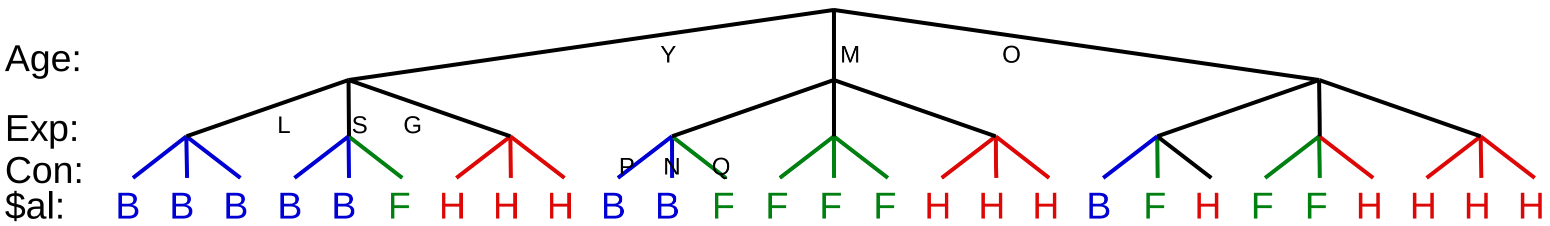



- Worst results (reversed order by input contribs)
- With 3 errors ➔ 7 rules:
  - (Or 4 errors ➔ 5 rules)

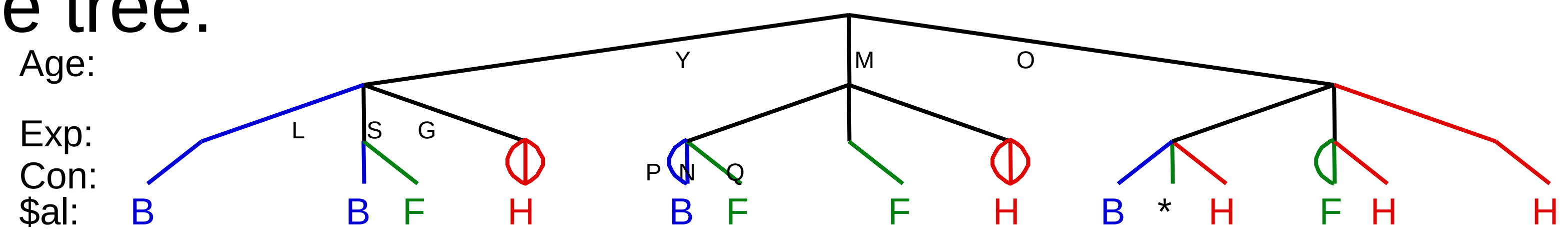


# Hierarchical dense rule bases

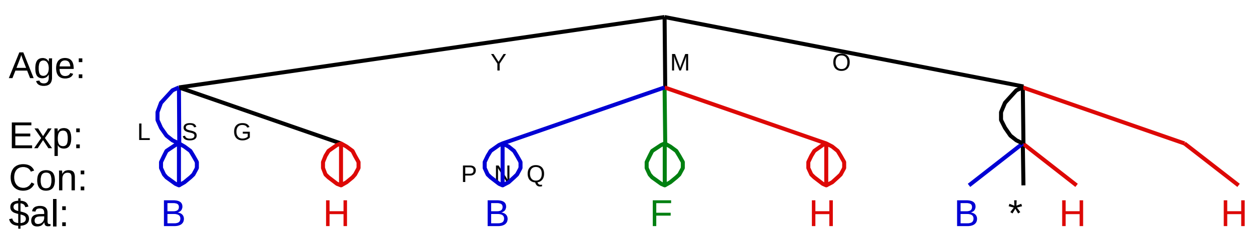
- Rules in tree (Age/Exp/Con) – different hier. seq.!




- Prune tree:

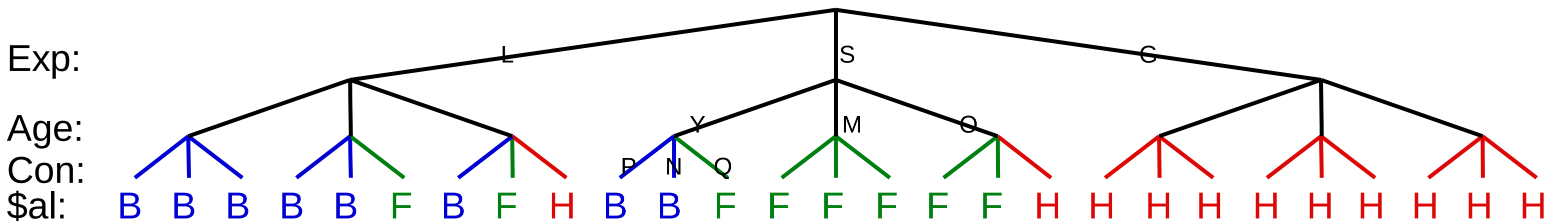



- Middling results, ignored input contributions
- With 3 errors
  ➔ 8 rules:

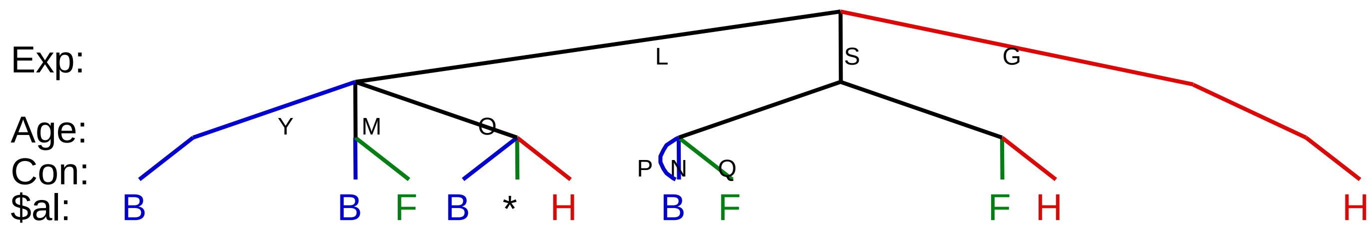


# Hierarchical dense rule bases

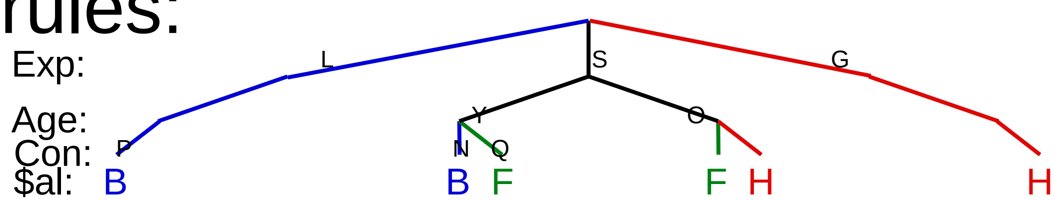
- Rules in a tree (Exp/Age/Con)




- Prune:



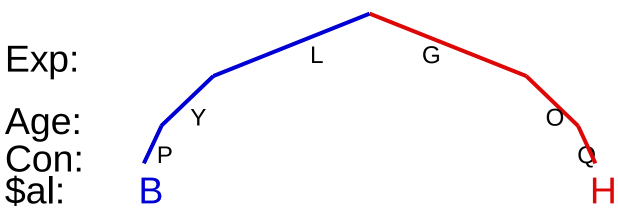

- Best results, uses decreasing input contributions
- Accept 3 errors ➔ 6 rules:
  - Interpolate between branches!
  - Performance now 89%



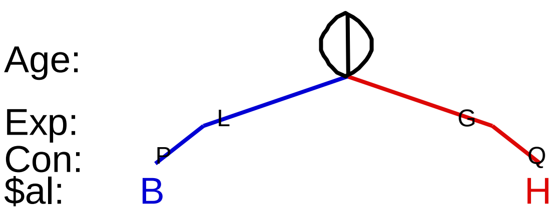
# Hier. dense rule bases

- How different are these really?

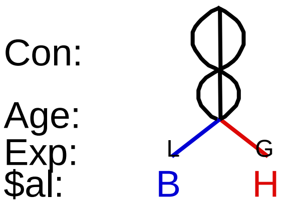
- Rules: Exp/Age/Con – 6 errors:
  - Performance is now 78%,
    only 2 rules, and using Exp only.




- Rules: Age only – 13 errors:
  - Result 52%, with 2 rules

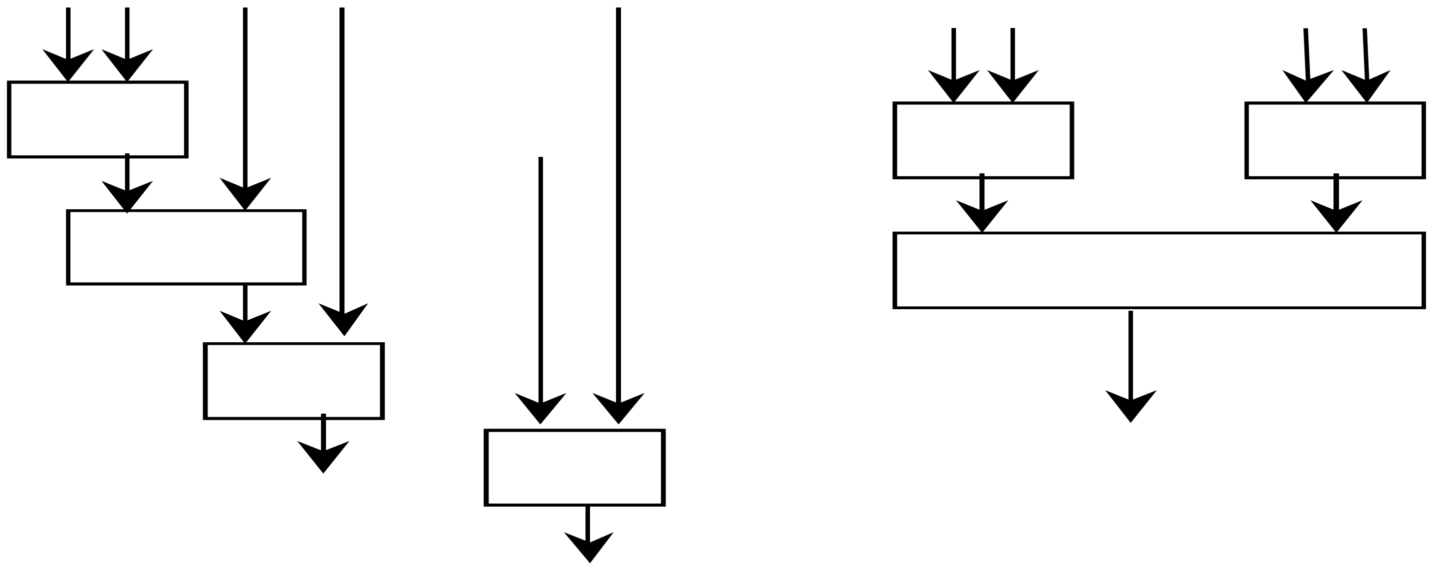



- Rules using Con only – 15 errors
  - Result 44%, with 2 rules



# Other hier. FZ models

- Advantage: Effective complexity reduction
- Disadvantage: Loss of interpretability
- Input passes through multiple levels of fuzzy system, each level modifies result based on some fuzzy rules.
- Transformation of input to output becomes hard to trace

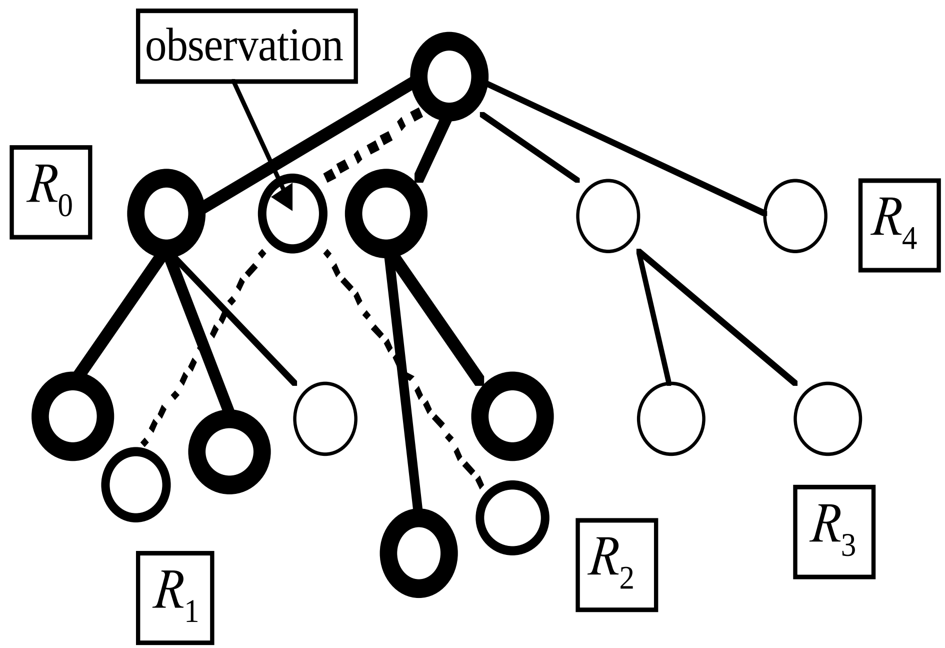


# Hierarchical rule bases

- Decompose multi-dimensional input state space

$$
\begin{aligned}
R0:\ & \text{If } z0 \text{ is } D1 \text{ then use } R1 \\
& \text{If } z0 \text{ is } D2 \text{ then use } R2 \\
& \vdots \\
& \text{If } z0 \text{ is } Dn \text{ then use } Rn
\end{aligned}
\qquad
\begin{aligned}
R1:\ & \text{If } z1 \text{ is } A11 \text{ then } y \text{ is } B11 \\
& \text{If } z1 \text{ is } A12 \text{ then } y \text{ is } B12 \\
& \vdots \\
& \text{If } z1 \text{ is } A1m_1 \text{ then } y \text{ is } B1m_1
\end{aligned}
$$

- Interpolate between branches



# Case Study

- Real world Petroleum Data
- The objective is to develop an estimator to predict porosity (PHI) from well logs.
- 8 Dimensional Inputs GR, RDEV, RMEV, RXO, RHOB, NPHI, PEF and DT
- 633 rows of data, same data used for training / testing
- Aim to construct a hierarchical fuzzy system with reasonable accuracy + good interpretability from real world data
- Lack of rule extraction techniques designed for hierarchical fuzzy rule base generation

# Case Study

- Convenient approach: Develop a conventional ('flat') fuzzy system, and then convert it to a hierarchical system.
- Brief description of Rule Extraction:
- Fuzzy cluster output space.
- For each output fuzzy cluster $B_i$
  - a cluster in the input space $A_i$ is induced.
- The input cluster is projected onto the various input dimensions to produce rules of the form:

  If $x_1$ is $A_{i1}$ and $x_2$ is $A_{i2}$ and … $x_n$ is $A_{in}$
  then y is $B_i$

# Case Study

- Conversion to Hier. Fuzzy Sys.:
  - Two or more fuzzy rules are merged to form hierarchical fuzzy rules. E.g., the two rules:

    If $x_1$ is $A_{11}$ and $x_2$ is $A_{12}$ then y is $B_1$
    If $x_1$ is $A_{21}$ and $x_2$ is $A_{22}$ then y is $B_2$

    can be merged to form:

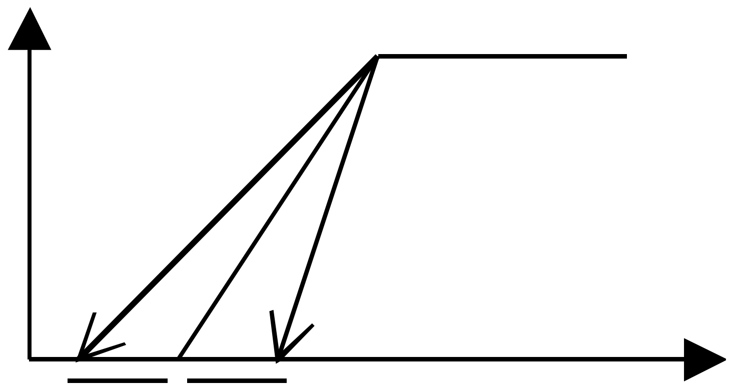
    If $x_1$ is ($A_{11}$ $A_{21}$) then use $R_1$
    $R_1$: if $x_2$ is $A_{12}$ then y is $B_1$
    if $x_2$ is $A_{22}$ then y is $B_2$

# Case Study

- Hier. version has more rules:
  - (1 meta rule + 2 rules) vs (2 rules)
- Inference more efficient in hierarchical version:
  - Number of terms in rule antecedents for the hierarchical version (3 terms) is less than the original version (4 terms).
  - For accuracy: $A_{11}$ and $A_{21}$ must coincide as much as possible – (by subjective evaluation).

# Hier. Fuzzy Modeling

- Perform parameter tuning to improve the performance of the hierarchical fuzzy system generated.

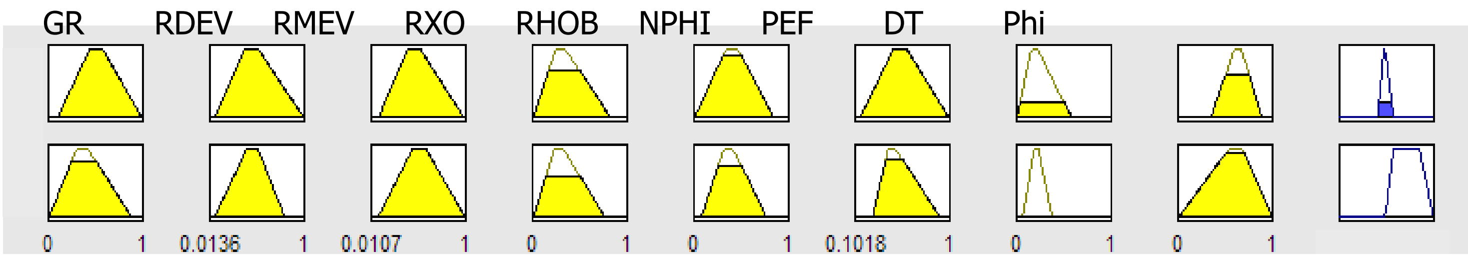


- Performance index:

$$PI = \sum_{i=1}^{m} (y^i - \hat{y}^i)^2 / m$$

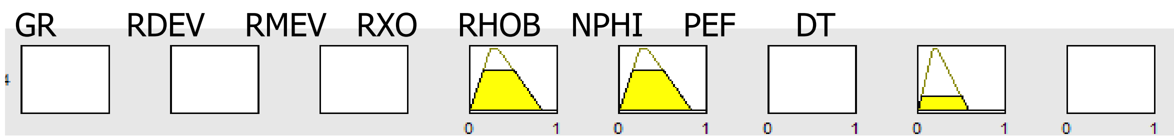
# Case Study

- Sample rules from original 'flat' fuzzy system




- Sample meta rule and its corresponding sub-rule base

**Meta Rule:**




Use R4

**R4:**