Mesure de la qualité d'un arbre

For example X_i : $f_{y_i}^{DT_j}(X_i)$ membership degree to its own class y_i with DT_j

Mesure de la qualité d'un arbre

For example X_i : $f_{y_i}^{DT_j}(X_i)$ membership degree to its own class y_i with DT_j

- Quality of a tree : depends on accuracy on difficult examples
  - examples taken in a reference set
  - degree of ease of an example to be correctly classified

$$\varepsilon(X_i) = \frac{\sum_{j=1}^{N_F} f_{y_i}^{DT_j}(X_i)}{N_F}$$

Mesure de la qualité d'un arbre

For example X_i : $f_{y_i}^{DT_j}(X_i)$ membership degree to its own class y_i with DT_j

- Quality of a tree : depends on accuracy on difficult examples
  - examples taken in a reference set
  - degree of ease of an example to be correctly classified

$$\varepsilon(X_i) = \frac{\sum_{j=1}^{N_F} f_{y_i}^{DT_j}(X_i)}{N_F}$$

- Degree of quality κ of a (fuzzy) decision tree
  - measures its ability to correctly classify difficult examples

$$\kappa(DT_j) = \frac{\sum_{i=1}^{n} (1 - \varepsilon(X_i)) \cdot cc_{ji}}{n}$$

with

$$cc_{ji} = \begin{cases} 1 & \text{if } X_i \text{ is correctly classified by } DT_j \\ 0 & \text{otherwise} \end{cases}$$

High Quality Trees

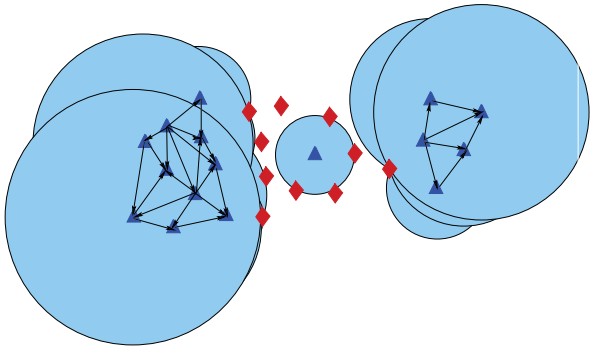
- Trees can be classed depending on their degree of quality
  - HQT : high quality tree DT_j : $\kappa(DT_j) \geq \kappa_t$
  - LQT : low quality tree
- Objective : to prune high quality trees
  - find sub-classes of high quality trees
  - each sub-class is represented by a single tree
  - Jason Forest's algorithm on $\{\text{HQT} \cup \text{LQT}\}$:
    - hypersphere classifier like

[Cooper 61, Reilly et al. 82, Marchand and Shawe-Taylor 02, Forest et al. 06]

  - idea : to highlight subgroups in a class

The Jason Forest's Algorithm : how it works

- Considering a training set



- Sub-classes of friends are highlighed
  - here 3 distinct sub-classes are found for the blue class

The High Quality Trees Forest approach

1. Selection of examples
   - training set / reference set / test set
2. Construction of a forest of fuzzy decision trees
   - random samplings of the training set
3. Evaluation of each tree on a reference set
   - computation of the degree of ease of each example
   - computation of the degree of quality of each tree
4. Pruning of the forest
   - discovery of sub-classes among high quality trees
   - selection of the representative high quality tree of each sub-class

$\Longrightarrow$ Size of the forest = number of sub-classes of HQT