

# Assignment: A3

---

Due: 4 October 2012

You are to explore the use of decision trees to classify scanned images of the 26 lower-case characters (i.e., a-z), and compare to the RBF method. Several aspects of decision trees deserve careful attention:

- **Input vector**: try at least 2 possibilities from these 3 alternatives:
  - Pixel values (binary): image pixels or subsets of image pixels
  - Low-level features: e.g., moments, projections
  - High-level features: Euler number, regions filled by letter, etc.
- **Entropy calculation**:
  - Describe entropy results at each step
  - Consider alternatives to entropy as a basis to select a feature
- **Data Management**:
  - How to select training and testing
  - How to compare learning methods

In addition, the results need to be presented in a strong statistical framework; this means computing statistics (e.g., mean, variance) over several trials (how many?), and showing confidence intervals.

Finally, the analysis and interpretation are the essential parts of the report; use these to present your findings, understanding and remaining problems.

In this assignment, the two major goals are to explore the use of decision trees and to compare the two learning methods (RBF and decision trees).

There is a set of sample images on the class data sub-directory.