

KNN Classifier Implementation

Project Proposal

Name of Classifier to Implement:

- The classifier K-Nearest Neighbor (KNN) will be implemented.

Programming Language Being Used:

- The programming language C++ will be used to implement this classifier

The Format of Training Data Being Used:

- Data set being used:
 - The training data corresponds to various measurements taken from three different varieties of wheat seeds: Kama, Rosa, and Canadian.
- Source of the data set:
 - The data used in this project was obtained from the UCI Machine Learning Repository.
 - URL: <https://archive.ics.uci.edu/ml/datasets/seeds>
- The attributes' types:
 - All attributes are real-valued continuous measurements of the wheat seeds.
- Number of attributes:
 - In total, there are seven different attributes.
- Attributes:
 - Area A
 - Perimeter P
 - Compactness C
 - Length of Kernel L
 - Width of Kernel W
 - Asymmetry coefficient AC
 - Length of kernel groove LG
- The size of the training set:
 - There are 210 total training instances.
 - 70 corresponding to Kama wheat seeds.
 - 70 corresponding to Rosa wheat seeds.
 - 70 corresponding to Canadian wheat seeds.