

Feature Selection Tool

Features

- Implements two common feature selection techniques:
 - Pointwise mutual information (PMI)
 - Pearson's chi-squared test (χ^2)
- Understands LIBSVM file format as an input
- Supports multi-threading

Feature ranking metrics

- In PMI mode the tool computes normalized PMI for each class and sum:

$$s_i = \sum_{c \in C} \left| \frac{PMI(f_i, c)}{\log p(f_i, c)} \right|$$

- In Chi^2 mode it constructs and solves contingency table using the basic formula:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Code organization and requirements

- There are four classes: one helper class for dataset assembly, one base class for evaluation algorithms and two concrete implementations
- Objective-C 2.0 and ARC must be supported by the compiler
- GCD will be used automatically if available in the system

