

Term Project: Feature Selection on Mars Crater Dataset

Educational Goal

Apply AI techniques to real-world Mars data.

Phase II: Informed Search on Feature Selection

(300 points)

Due: 4:00 PM Monday December 13, 2010

Requirements

- **Supervised Learning:** Use an **informed search algorithm of your choice** to find the best feature subset out of the total 1089 features that can achieve the highest F1 score on test set TestSet3.csv using training set train.csv.
- Use Weka LibSVM to build a classifier from the training set and use the built classifier to classify the test set, using the feature subsets you have selected.
- **Evaluation function of the informed search algorithm.** Discuss the heuristic function you have designed: is it admissible or consistent? Prove it.
- **Property of the informed search algorithm.** Is it complete? Is it optimal? What is the time complexity of the algorithm? What is the space complexity of the algorithm? Prove your conclusions.
- **Informed search algorithm.** Draw a flowchart of your informed search algorithm.
- Write an experiment report to discuss your experimental results, including detailed parameter settings and experimental results.
- **Note : Please do all the tasks of classification, feature selection and search algorithm in ONE program. Demonstrate all the functions of the project in one run. Do not divide those tasks into separate programs.**

Submission Requirements

1. Prepare a readme file for your TA to run your project on her machine. Your program should be well-documented. Variable names and function names should be self-descriptive. Major functions should be explained clearly .The program outputs should be clearly presented.
2. Submit the paper copy of the report, and source code of the scripts with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.

3. Submit the softcopy of the report and scripts through your UMassOnline account at <http://boston.umassonline.net/index.cfm>.
4. Zip all the files. One submission per team. Save the file as `sdm_teamNumber`. For example, Team 1 should name their file as `sdm_team1.zip`.
5. No hard copies or soft copies results in 0 points.