Predictive Analysis on Optimization of Cancer Classification by Machine Learning Methods

Lei Liu
Sun Microsystems, Inc.
Prepared for Rocky'07
Introduction

• Classification Problem
  > Class Discovery
  > Class Prediction
  > Data Structure
  > Algorithms
  > Applications of Classification
    > Cancer Classification
    > Molecular Classification Report
    > The Published Datasets
Problem

- **Cancer Classification**
  - *Small Samples*
  - *High Dimensional State Spaces*
  - *Learning*
  - *Class Prediction*
  - *Class Discovery*
  - *Optimization*
  - *Dataset Processing*
Solution

- **Parametric Machine Learning Methods**
  - Attribute Selection
  - Class Prediction
  - Data Structure
  - Algorithms
  - Machine Learning Methods

- **Performance Evaluation**
  - Cross Validation
  - AUC
Recommendation

• **Exposing Parallelism**
  > **Hardware**
    > CMT, Sun T2000 Platforms
  > **Software**
    > *Fortress Programming Language*
  > **Algorithms**
    > *Data Distribution, Reduction, Aggregation*
  > **Evaluation**
    > *Speedup, Parallel Stack*

• **Data Management**
  > Sun HPC and Virtual Tape Storage
  > Policy-based Replication