CLASSIFICATION ANALYSIS OF COCOMO 2.0 DATA

Richard W. Selby
Dept. of Information & Computer Science
University of California
Irvine, California 92697

714-824-6326
714-725-6411, fax
selby@ics.uci.edu
Classification Analysis Techniques

- Scalable to large projects
- Calibratable to new environments
- Measurements are integratable
- Leverage previous experience
- Process, product & team attributes

Focus on High-Payoff Areas: the 80:20 rule

Example:

*+*: Classified as likely to have property P (e.g., integration errors)

*-*: Classified as unlikely to have property P
Amadeus Project

Consistency vs. Completeness

- 16 NASA systems
- Target: Identify fault-prone (top 25%) and cost-prone (top 25%) components
- 960 configurations

Legend:
- Consistency opt.
- No optimization
- Completeness opt.

Consistency
(= 100 - percent false positives)

Completeness (= 100 - percent false negatives)
PRELIMINARY RESULTS: COCOMO 2.0 DATA

- Classification tree generation tool calibrated to COCOMO 2.0 data
- Analyzed 66 COCOMO 2.0 data points
- Used lower and upper bounds for effort and duration to define “target classes” of projects
- Classification trees can integrate both symbolic values (such as “Low”, “Very high”) and numeric values
- High variance in prediction results due to small sample size
- Analysis at the module-level would enable larger sample sizes and finer grain predictions