Obtaining the Cost

How do we determine cost?

The effectiveness of CSL relies on the supplied cost matrix.

- Low cost will not find the proper classification boundary
  - ➤ biased towards majority class
- High cost may impair generalization
  - ➤ Unable to predict well the majority classes

How do we determine cost?

Cost Matrix provided by Expert

Domain expert to determine real costs

Heuristic

Imbalance Ratio

Optimisation

The nature of imbalanced classes

Factors that influence the ability of a classifier to identify rare events

- Small sample size
- Class separability
- Within-class sub-clusters.

THANK YOU

www.trainindata.com