Data Mining
(Section 19.2)

Peter Revesz

CSCE 413/813
Computer Science and Engineering
University of Nebraska – Lincoln

Data Mining Example

A headhunting company needs to find an artistic and financially reliable applicant for director of an art foundation.

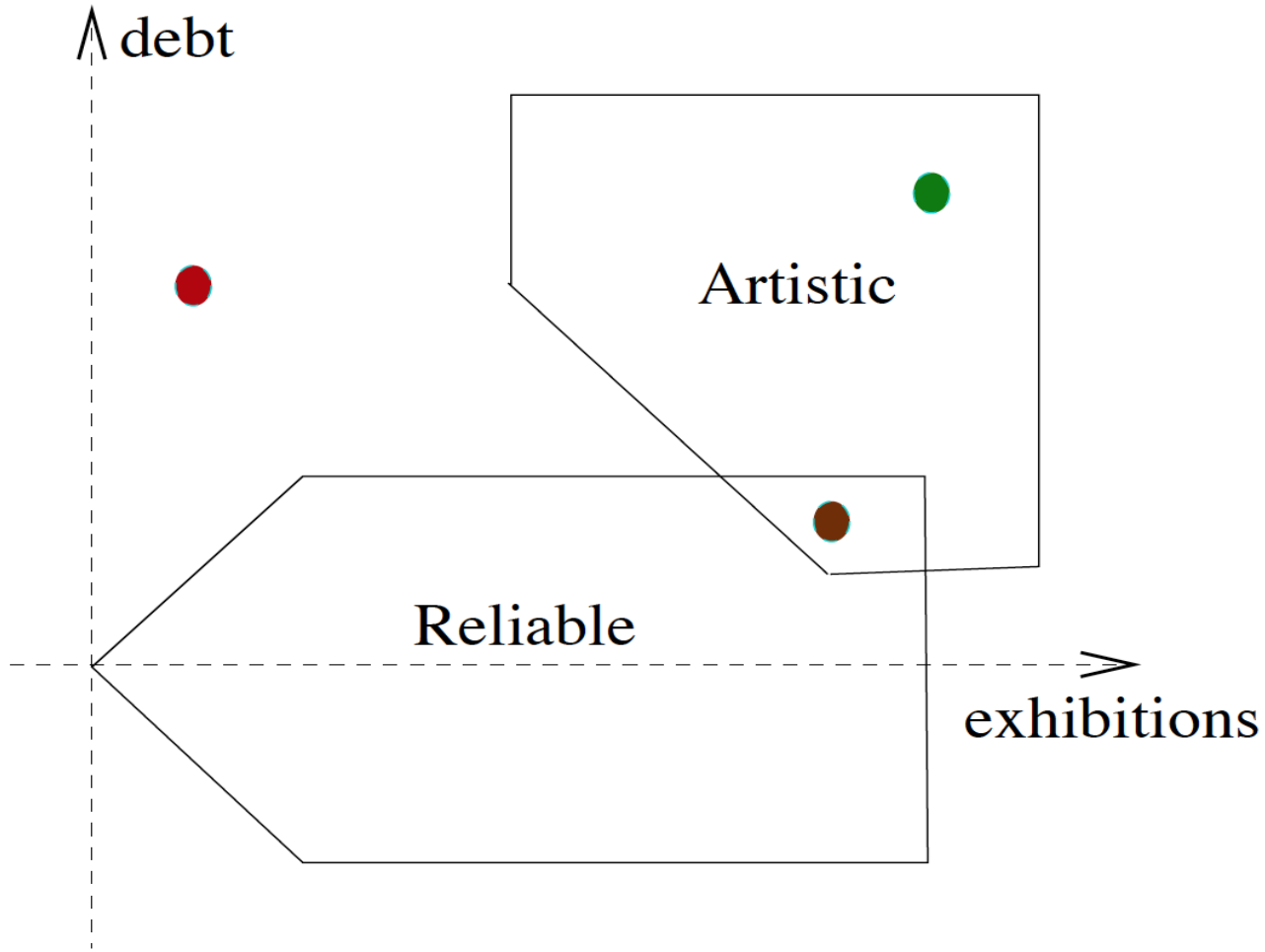
Applicant

Name	Exhibitions	Debt
Brown	70	15
Green	80	50
Red	10	40
⋮	⋮	⋮

Exhibitions = number of exhibitions made by the applicant.
Debt = amount of personal debt in thousands of dollars.

Data Mining

Based on previous experience, a decision tree or a SVM can be used to identify a region of the feature space, which describes artistic and reliable.



Data Mining

The feature space regions can be represented by a constraint database.

Personality

Type	Exhibitions	Debt	
t	x	y	t = "Artistic", $x \geq 40, x \leq 90, y \geq 10, y \leq 60, y \geq -x + 80$
t	x	y	t = "Reliable", $x \leq 80, y \geq -20, y \leq 20, y \leq x, y \geq -x$

Data Mining

The artistic and reliable applicants can be selected by a SQL query:

```
CREATE VIEW   Promising_Candidate(Name)
SELECT        Name
FROM          Applicant AS A,  Personality AS P
WHERE         A.Exhibition = P.Exhibition  AND
              A.Debt = P.Debt  AND
              P.Type = "Artistic"
INTERSECT
SELECT        Name
FROM          Applicant AS A,  Personality AS P
WHERE         A.Exhibition = P.Exhibition  AND
              A.Debt = P.Debt  AND
              P.Type = "Reliable"
```

Data Mining

Challenge Question: List all the data mining tasks that the headhunting company is paid for.

Data Mining

Challenge Question: List all the data mining tasks that the headhunting company is paid for.

A. Creating a relational database for the job applicants.

B. Doing previous studies using decision trees or SVMs.

C. The previous studies likely included data integration.

D. Creating a constraint database for the relevant feature space regions.

E. Writing SQL queries for the identification of promising candidates.

F. Potentially also decision support in further ranking the candidates.