sergey@cs.stanford.edu
Oct 13 1998

CS 349

Classification and Singular Value Decomposition

Sergey Brin

October 13, 1998

Stars, Galaxies, and Nebulas

Input variables: 25 x 25 pixel images

- **625 dimensional vectors**

Output:

- **star, galaxy, cluster, ..., other**

Outline

Classification Techniques

- Decision Trees
- Neural Nets

Vector Models

- Singular Value Decomposition

Classification Applications

Astronomy

Credit Card Applications

Credit Card Fraud

Submarines

Web Pages

Types of Craters

sergey@cs.stanford.edu

Basic Classification Techniques

Train on training data

- input tuple, output value
- input tuple, output value
- input tuple, output value

Process Training Data (Learning)

- build data model (technique dependent)

Classify new data

- input tuple -> predicted output
- input tuple -> predicted output

Common Classification Techniques

Constant - always output most common class

Decision Tables - entry for every possible input

Decision Trees - C4.5

Basic Decision Trees

Example decision tree:

Generating Decision Trees

Try every attribute.

Pick the one that is best predictor of output.

- **(needs to be defined)**

Split based on that attribute.

Repeat for every branch.

Break

5 minutes

Singular Value Decomposition

Motivation

Graphical Example

High Dimensional Vector Datasets

Examples:

- **Word Occurrence**
- **Time Series**

Hard to handle

Need Lower Dimensional Representation

Example 2-dimensional dataset

Can we find a good 1-dimensional representation (projection)?

Picking a lower dimensional space

Covariance Matrix

Eigenvectors

sergey@cs.stanford.edu

Other Matrices:

A*A for word occurrence (LSI)

Link Graph: G

G*G

Summary:

Classification

Singular Value Decomposition