Principle of Locality for Statistical Shape Analysis

Paul Yushkevich

Outline

- Classification
- Shape variability with components
- Principle of Locality
- Feature Selection for Classification
- Inter-scale residuals

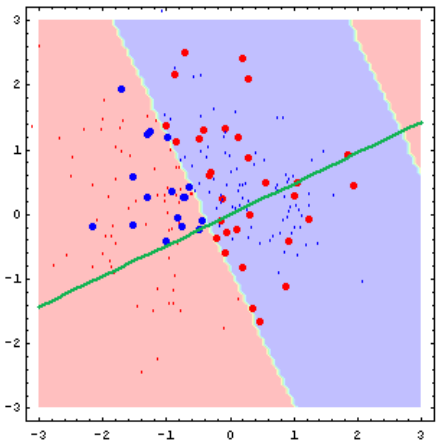
Classification

- Given training data belonging to 2 or more classes, construct a *classifier.*
- A classifier assigns a class label to each point in space.
- Classifiers can be:
  - Parametric vs. Nonparametric
  - ML vs. MAP
  - Linear vs. Non-linear

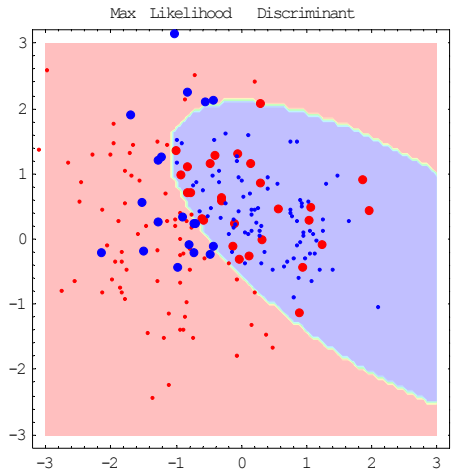
Common Classifiers

- Fisher Linear Discriminant
  - Linear, Parametric (Gaussian model with pooled covariance)
- K Nearest Neighbor, Parzen Windows
  - Non-linear, non-parametric
- Support Vector Machines
  - Non-parametric, linear

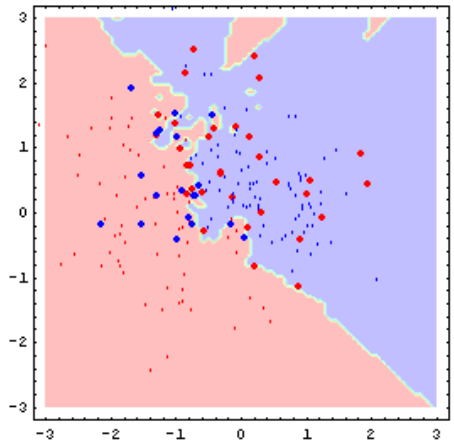
Examples



Fisher Classifier



Non-linear MLE Classifier

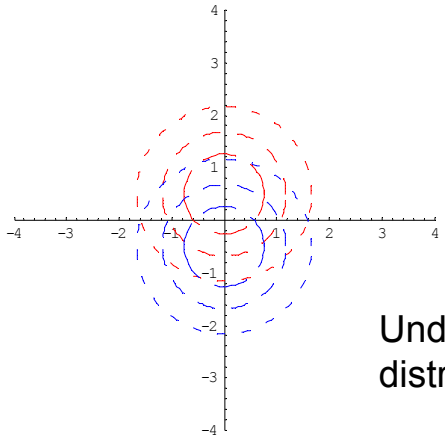


3-Nearest Neighbor

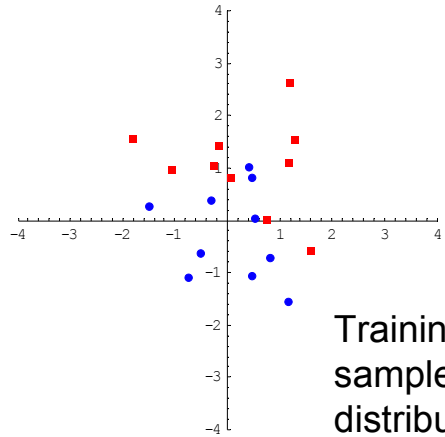
Feature selection

- If we limit classification to just the right subset of features, we can drastically improve the results

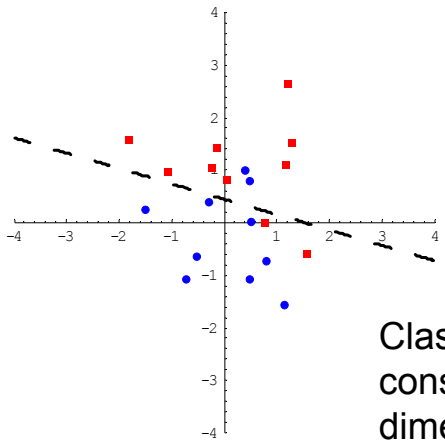
Feature Selection Example



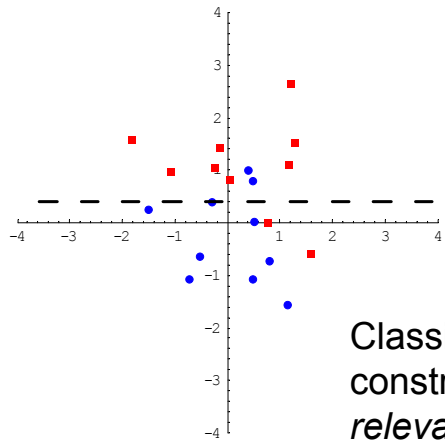
Underlying distributions



Training Set sampled from the distributions

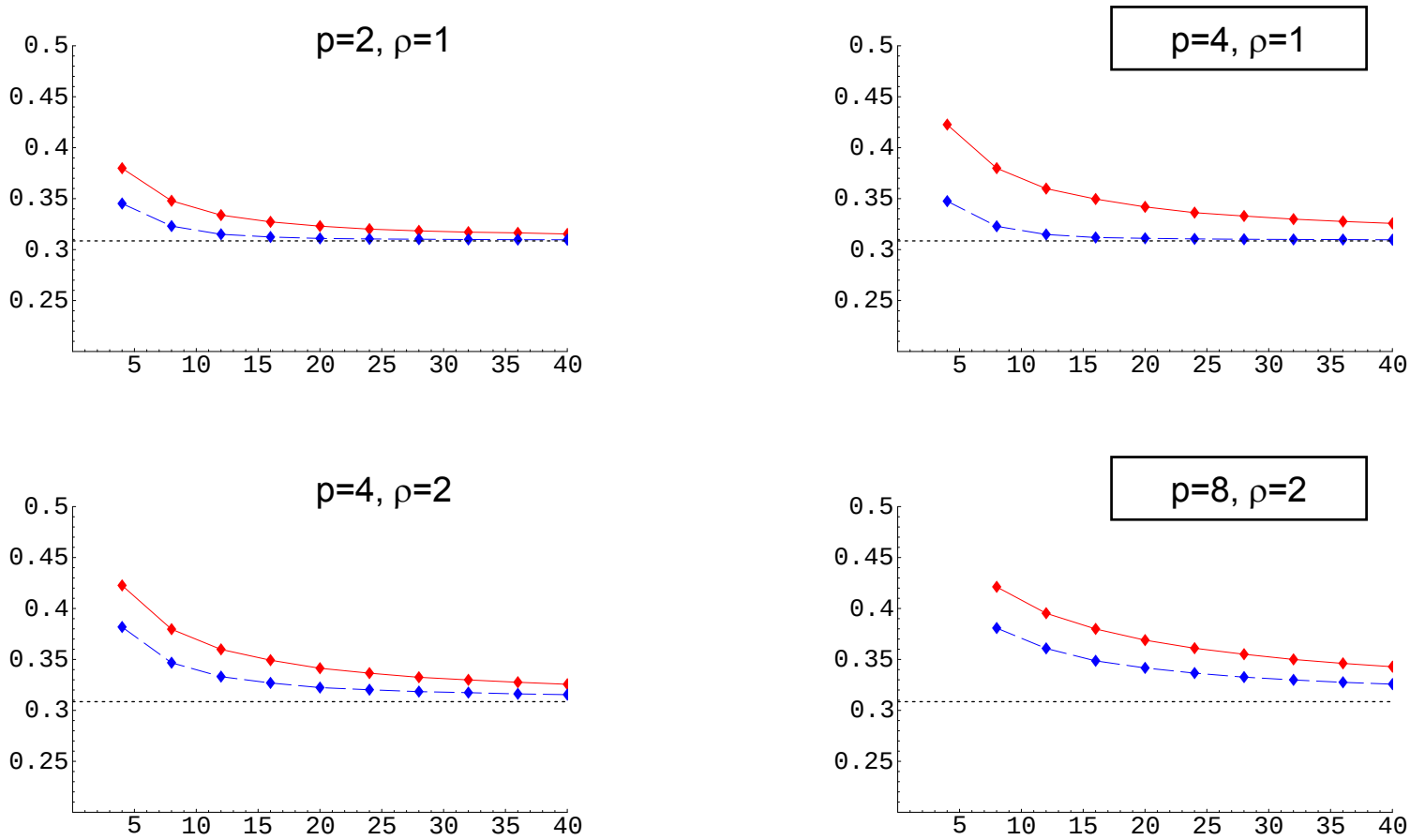


Classifier constructed in 2 dimensions



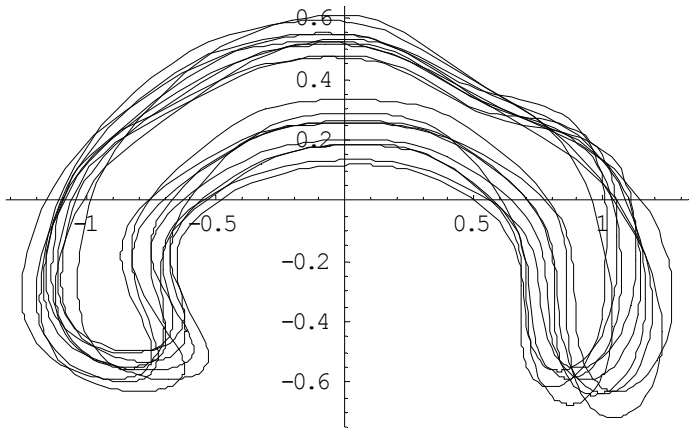
Classifier constructed in the *relevant* *y*-dimension

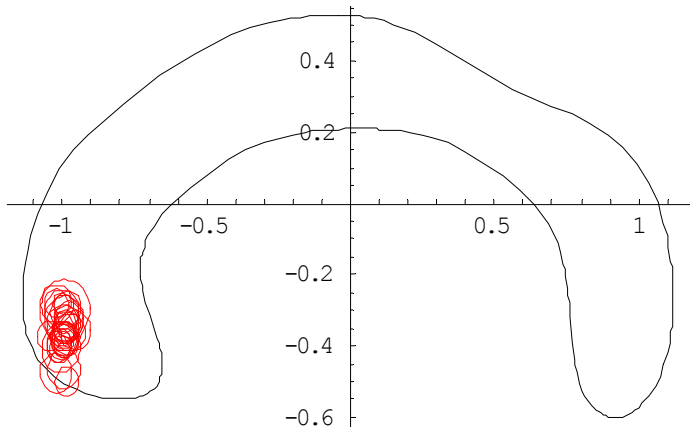
Feature Selection Example

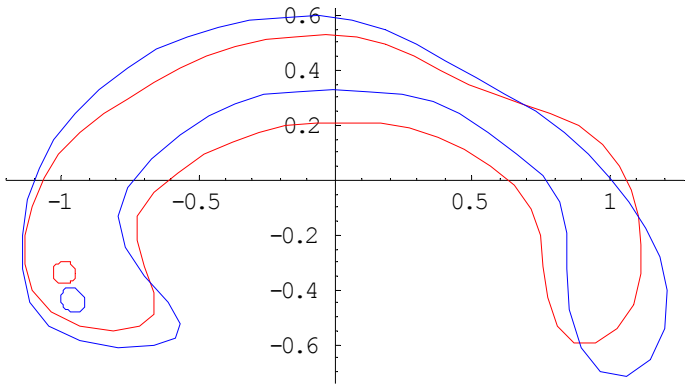


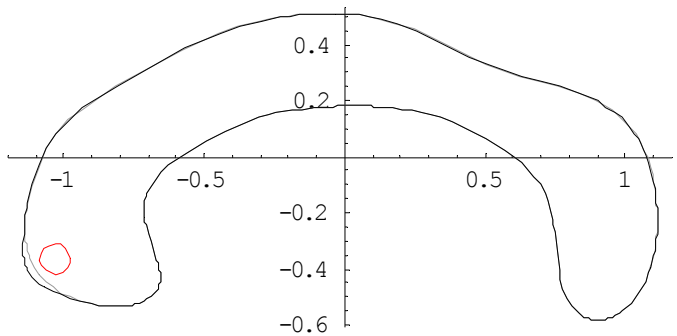
Expected error rate in classification in p dimensions vs. classification in the relevant ρ dimensions, plotted against training sample size.

Synthetic Shapes



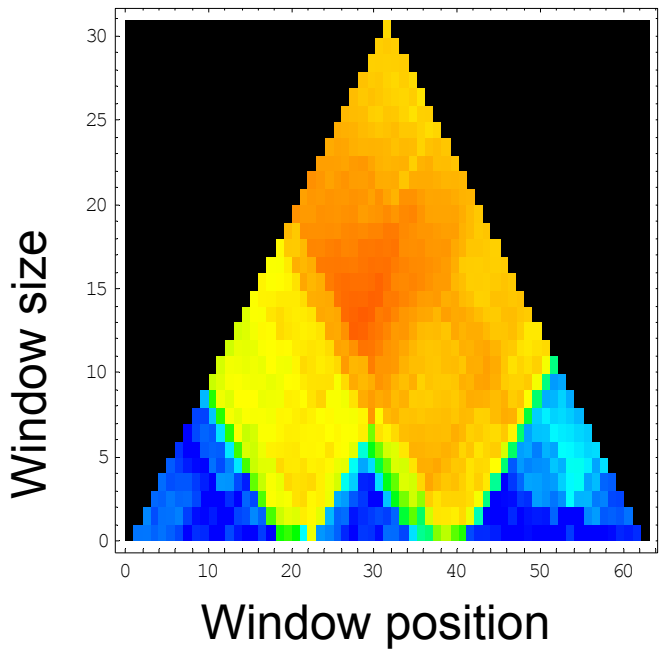






Feature Selection in Shape

- Relevant features should form a small set ‘windows’
- The order of features is important
- Example:
  - Fisher classification rate over all possible windows



Feature Search Algorithm

- Look for best set of feature windows and best classifier simultaneously
- Similar to SVMs
- Minimize:

$$E_{separation} + E_{Nfeature} + E_{Nwindows}$$