The Multilevel Classification Problem and a Monotonicity Hint

Malik Magdon-Ismail (RPI)
Hung-Ching (Justin) Chen (RPI)
Yaser S. Abu-Mostafa (CALTECH)

August 13 2002

➩ **MOTIVATION**

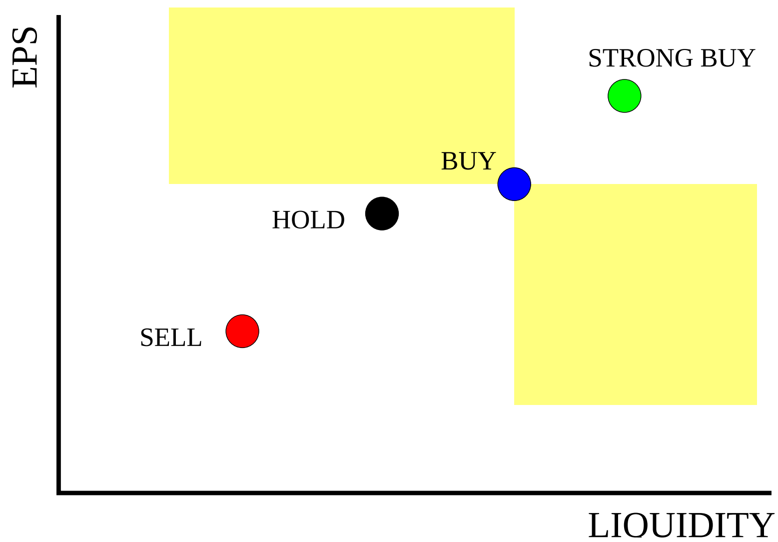
MULTILEVEL CLASSIFICATION

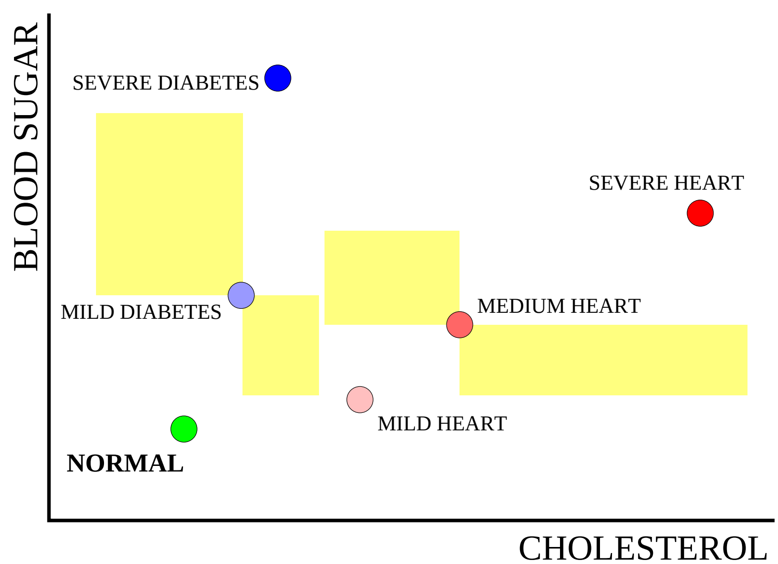
MONOTONICITY

ALGORITHMS

EXPERIMENTS

Motivation





Stock Selection, Disease Prediction, Fault Diagnosis, Weather Patterns,. . .

MOTIVATION

➾ **MULTILEVEL CLASSIFICATION**

MONOTONICITY

ALGORITHMS

EXPERIMENTS

Problem Setup

(K,L)-Multiclass-Multilevel Problem

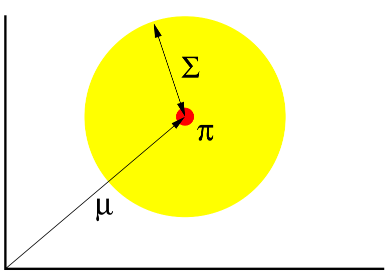
K categories, $c_1 \ldots c_K$.
L levels within each category, $l_1, \ldots, l_L$.

$$\text{Total \# of classes} \sim K \times L$$

Feature **dimension d**: $\mathbf{x} \in R^d$.

$N_{c,l}$ **data points** per category, level pair.

Gaussian Class Conditional Density



Class Conditional	True	Sample
mean	$\mu_{c,l}$	$\mathbf{m}_{c,l}$
covariance	$\mathbf{\Sigma}_{c,l}$	$\mathbf{S}_{c,l}$
prior	$\pi_{c,l}$	$\widehat{p}_{c,l}$

Bayes Optimal Decision Rule

Given $\mathbf{m}$, $\mathbf{S}$, $\mathbf{p}$ for all classes, and a *risk matrix* $\mathcal{R}$, one can compute the **Bayes optimal decision rule**

$$\begin{aligned} \alpha(\mathbf{x}) &= \operatorname*{argmin}_{c,l} \mathcal{R}(c,l|\mathbf{x}) \\ \mathcal{R}(c,l|\mathbf{x}) &= \sum_{c',l'} R(c,l|c',l')P(c',l'|x) \end{aligned}$$

MOTIVATION

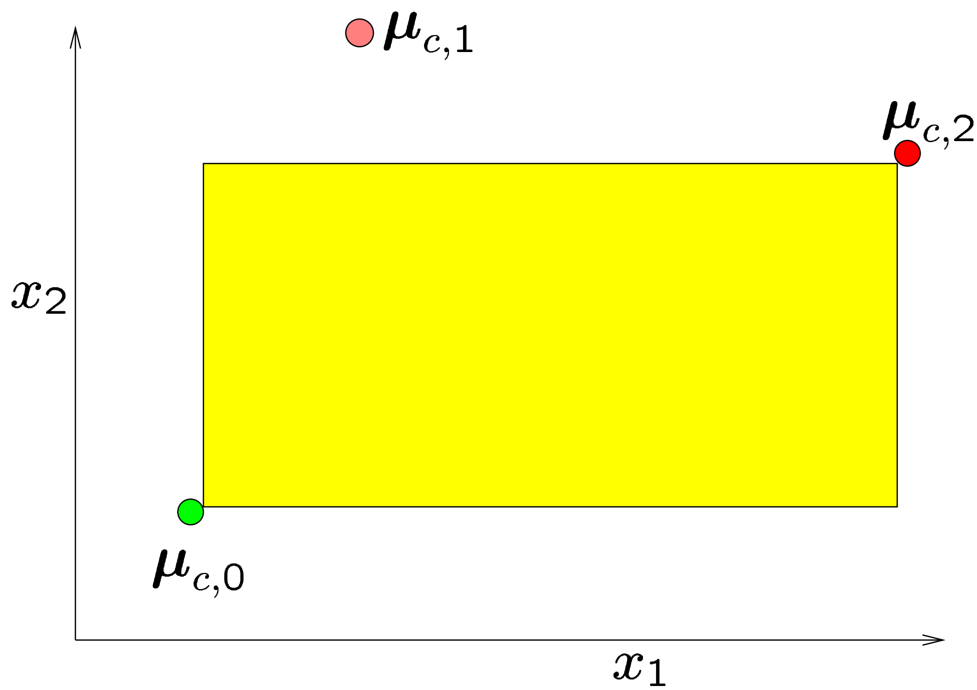
MULTILEVEL CLASSIFICATION

➾ **MONOTONICITY**

ALGORITHMS

EXPERIMENTS

Monotonicity



$$(\boldsymbol{\mu}_{c,1} - \boldsymbol{\mu}_{c,0}) \,.\!^*\, (\boldsymbol{\mu}_{c,2} - \boldsymbol{\mu}_{c,1}) \tag{1}$$

Monotononicity should hold within **every category** for **every triple** $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3$.

Incorporating Monotonicity

The sample means may **not** be monotonic.

Let $\hat{\mathbf{m}}_{c,l}$ be new estimates for the means. One can compute the **likelihood**

$$P\left(\{\mathbf{m}_{c,l}\}|\{\hat{\mathbf{m}}_{c,l}\}\right) \tag{2}$$

Use monotonicity to guide the choice of **prior**.

$$prior = \begin{cases} 1 & \text{(monotonicity satisfied),} \\ 0 & \text{(otherwise).} \end{cases}$$

MOTIVATION

MULTILEVEL CLASSIFICATION

MONOTONICITY

➯ **ALGORITHMS**

EXPERIMENTS

Optimization Problem

Minimize with respect to $\{\widehat{\mathbf{m}}_{c,l}\}$

$$\frac{1}{2}\sum_{l,c} N_{c,l}(\widehat{\mathbf{m}}_{c,l}-\mathbf{m}_{c,l})'\widehat{\mathbf{S}}_{c,l}^{-1}(\widehat{\mathbf{m}}_{c,l}-\mathbf{m}_{c,l})+\frac{1}{2}\log|\widehat{\mathbf{S}}_{c,l}|$$

with non-linear inequality constraints

$$(\boldsymbol{\mu}_{c,j}-\boldsymbol{\mu}_{c,i}).^*(\boldsymbol{\mu}_{c,k}-\boldsymbol{\mu}_{c,j}) \geq \mathbf{0}, \qquad \begin{array}{l} 1 \leq c \leq K \\ 0 \leq i < j < k \leq l_c \end{array}$$

$\widehat{\mathbf{m}}_{c,l}$ =estimated mean.
$\widehat{\mathbf{S}}_{c,l}$ =estimated covariance matrix.

Optimization Algorithms

Monte Carlo

Generate **random monotonic solutions** and descend on the objective function until −gradient has no component in a monotonic direction.

Keep the **best** solution.

Gradient Based Approach

Construct an **error function**

$$E(\hat{\mathbf{m}}_{c,l}) = E_{like} + \Omega E_{mon} \quad (3)$$

E_{mon} is designed to have a minimum when monotonic.

MOTIVATION

MULTILEVEL CLASSIFICATION

MONOTONICITY

ALGORITHMS

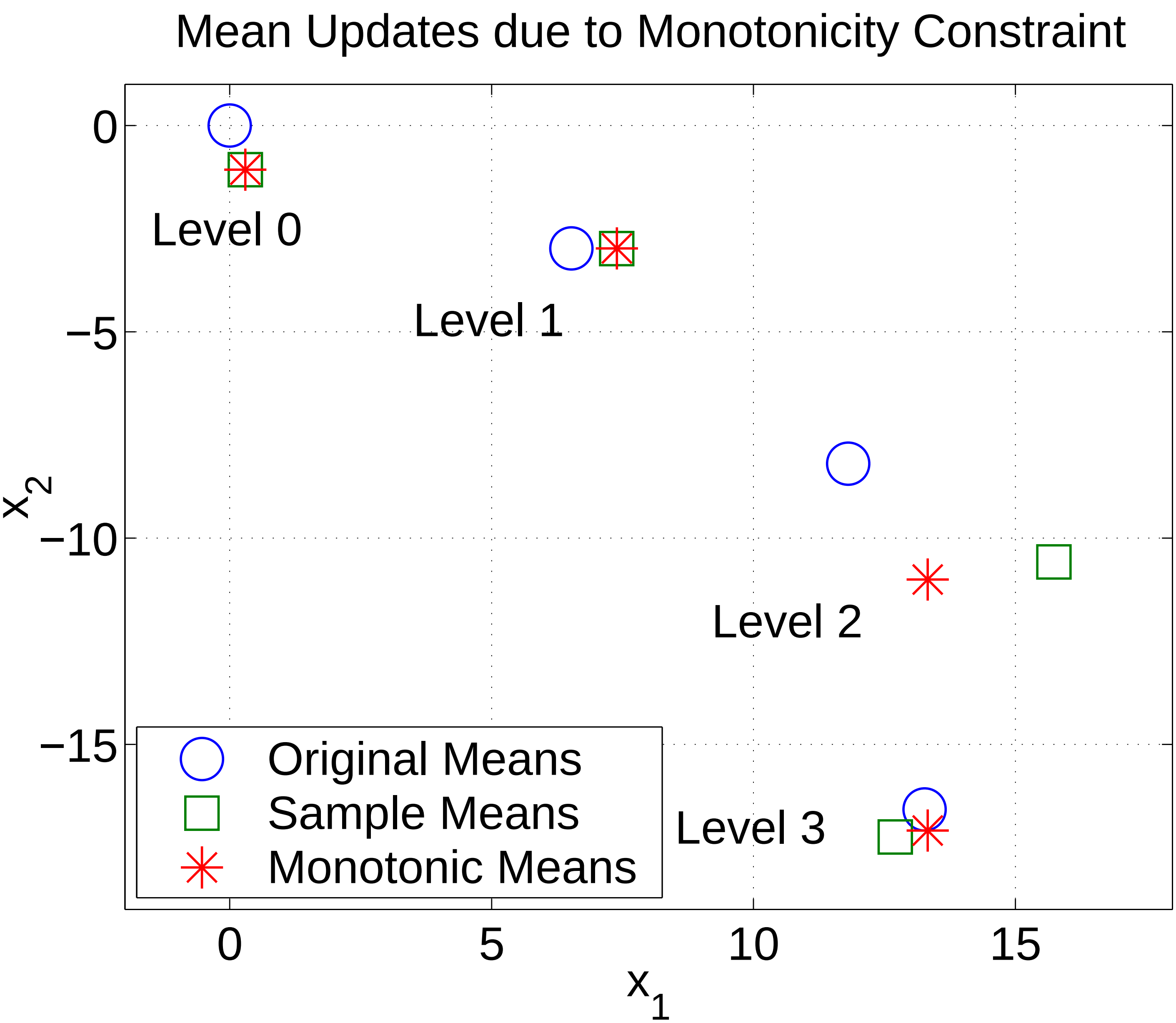
➩ **EXPERIMENTS**

Data Sets

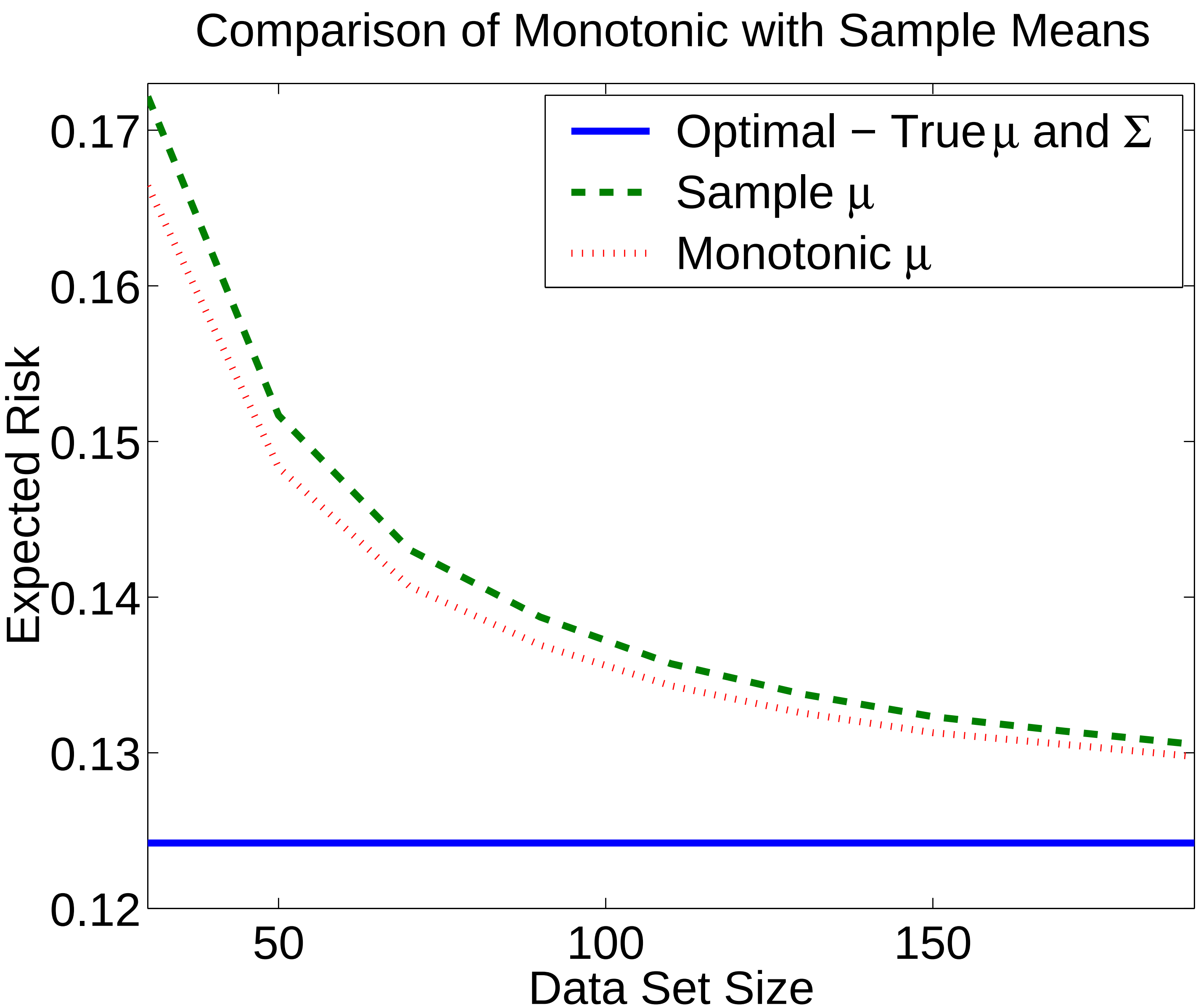
1. Toy data set created by monotonic teacher. (3 category)

2. UCI heart disease data. (1 category)
   – M/F split can make it 2 categories.

Mean Update Example

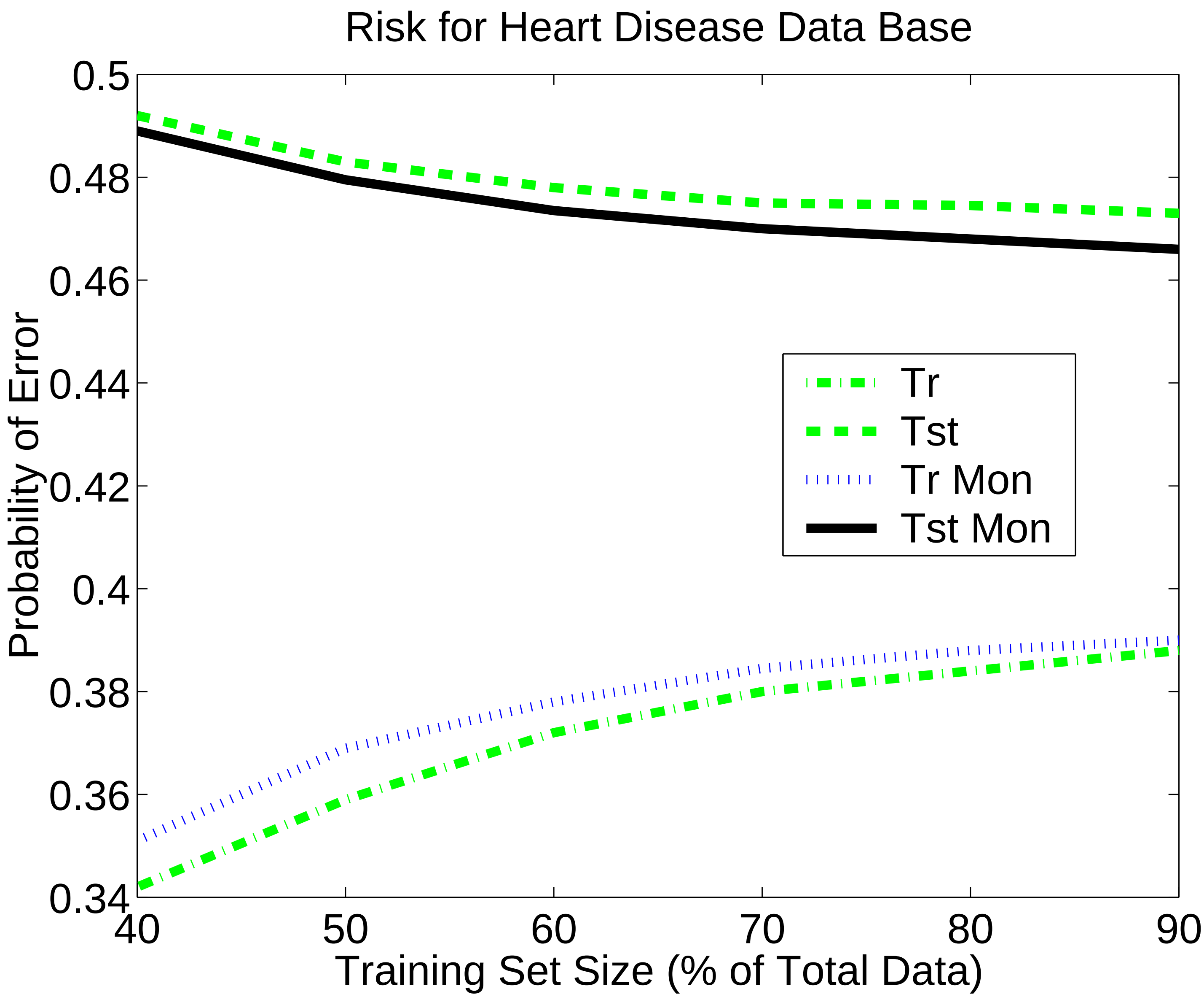


Toy Problem



Comparison of Monotonic with Sample Means
Optimal − True μ and Σ
Sample μ
Monotonic μ
0.17
0.16
0.15
0.14
0.13
0.12
Expected Risk
50
100
150
Data Set Size

Heart Data


Risk for Heart Disease Data Base
Probability of Error
0.5
0.48
0.46
0.44
0.42
0.4
0.38
0.36
0.34
40
50
60
70
80
90
Training Set Size (% of Total Data)
Tr
Tst
Tr Mon
Tst Mon

Closing

1. Multilevel property adds monotonic structure.

2. Montecarlo (slower) and Gradient based (fast) algorithms.

3. Results on real data promising.

Extend beyond Gaussian class conditionals?