



クラスタリングに対する 例からの学習

神嶌 敏弘 新田克己

電子技術総合研究所 推論研究室



本研究の目的

クラスタリングを対象にした新たな学習
問題「クラスタリングに対する例からの学
習」の定式化

この問題に対する学習方法の提案

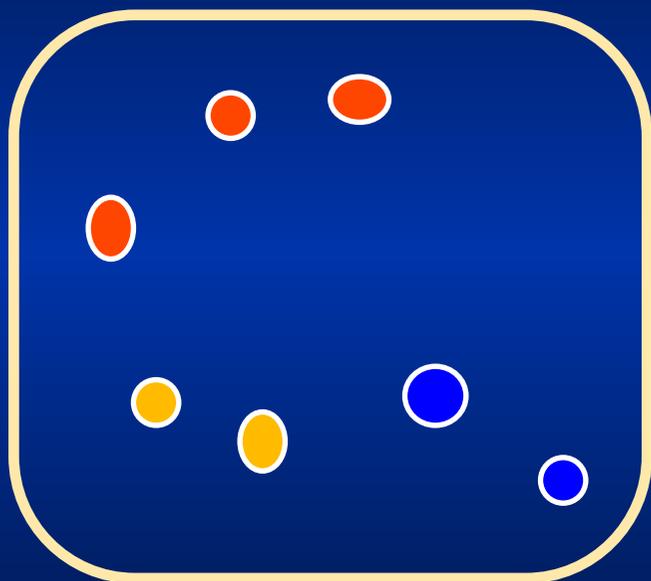
実験結果に対する定量的な評価をもと
に、この学習問題の特徴について考察

目次

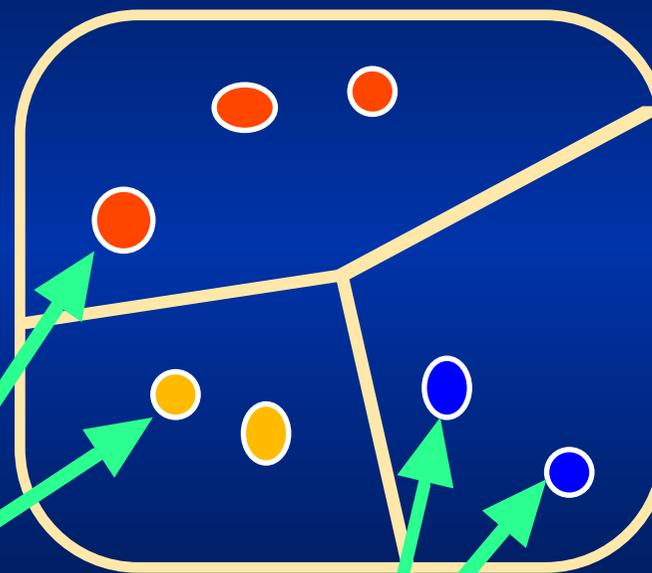
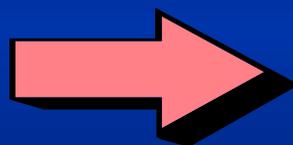
- クラスタリングに対する例からの学習とは
 - クラスタリングを応用する場合の問題点
 - この問題を解決するためにクラスタリングに対する例からの学習を定式化
- 学習と評価の方法
 - クラスタリングに対する例からの学習の方法
 - 学習結果の定量的な評価
- 実験と考察
 - 定量的評価と定性的評価との対応
 - 本学習と一般の「例からの学習」との相違

クラスタリング

分類対象の集合



分類対象の集合を
クラスタに分割



似ていない

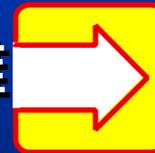
似ている

クラスタリングを 応用する場合の問題点

クラスタリング

- ・ 似たものを集めた分割を獲得する方法

対応付けが困難



学習を利用

- ・ ある望ましい分割を自動的に獲得することを要求

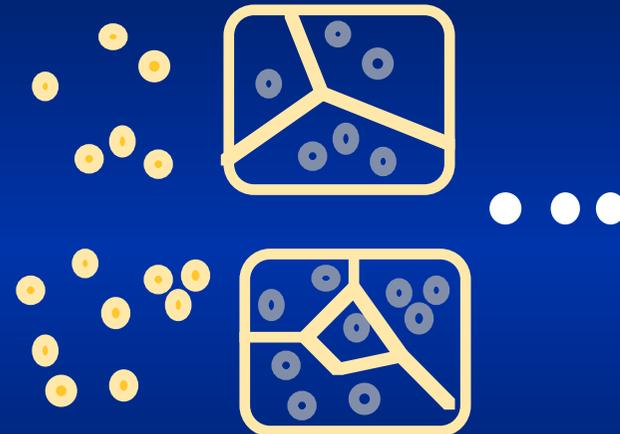
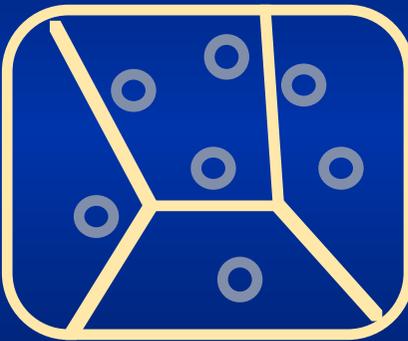
クラスタリングの応用分野

画像処理の領域分割など

クラスタリングに対する 例からの学習

学習事例

分類対象の
集合

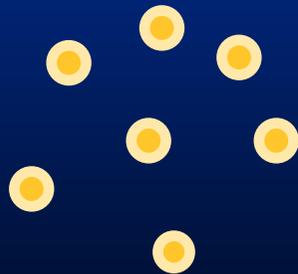


分類対象の集合を分割する基準

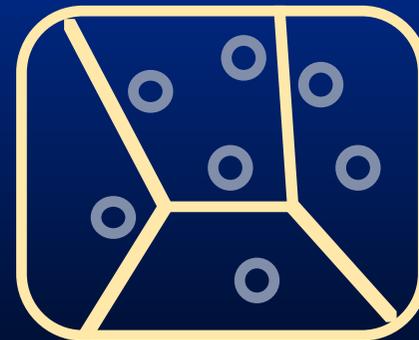
クラスタリングに対する 例からの学習（続き）

学習により獲得した
分類対象の集合を分割する基準

未知の
分類対象の集合



望ましい分割



クラスタリングに対する 例からの学習の方法

学習：分割を獲得するための基準を獲得

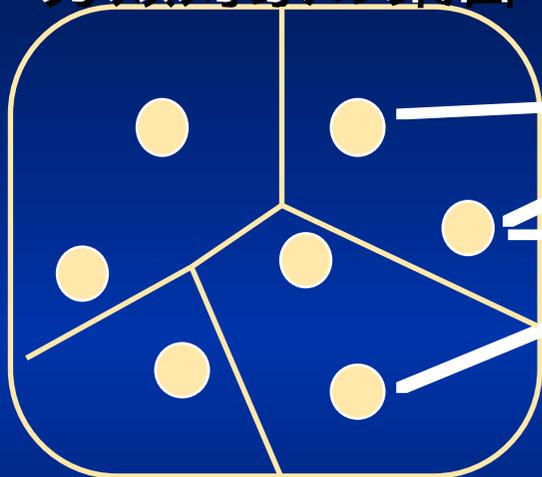
一対の分類対象が同じクラスタの要素になるかを
判別する規則を獲得

学習結果の適用：獲得した基準を利用して
分割を獲得

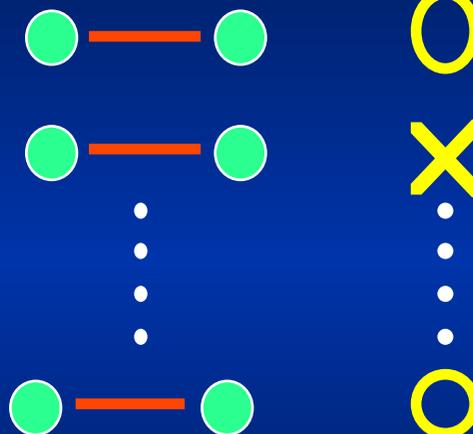
1. 分類対象の対を判別
2. 判別結果を利用した分割の獲得

分割の獲得のための学習

学習事例中の
分類対象の集合



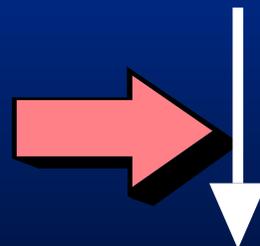
適宜選択された
分類対象の対



分類対象の対が
同じクラスタの
要素？



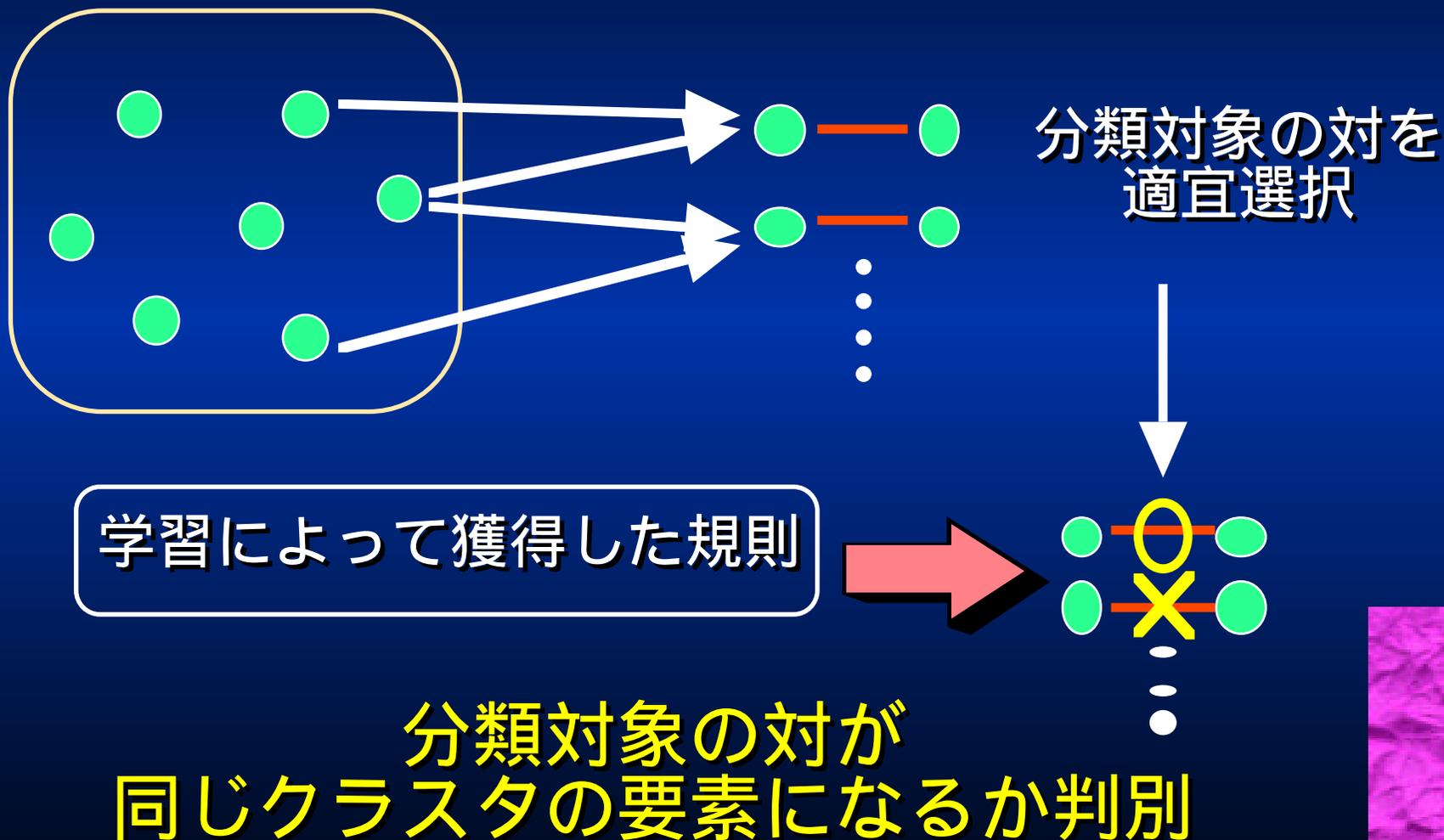
一般の例からの学習手法



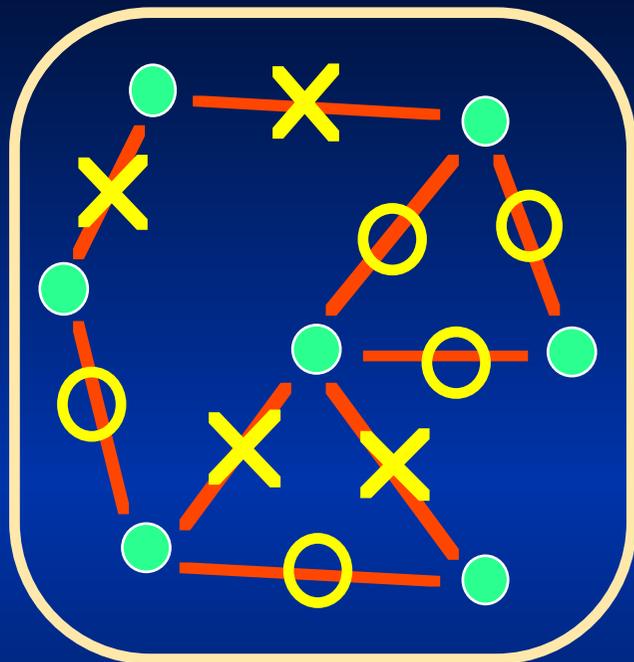
分類対象の対が同じクラスタの
要素になるか判別する規則

分類対象の対の判別

未知の分類対象の集合



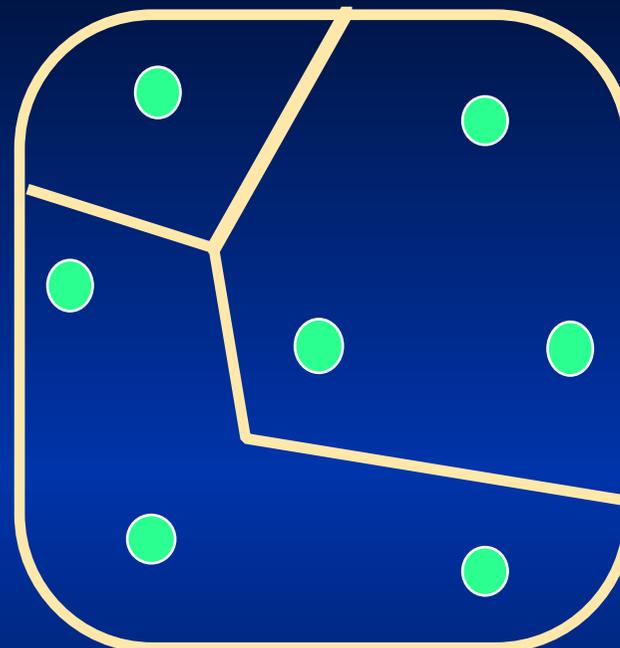
判別結果利用した分割の獲得



分類対象の対の
判別が終了した
分類対象の集合

クラスタ分析手法

最小距離法
最大距離法
群平均法



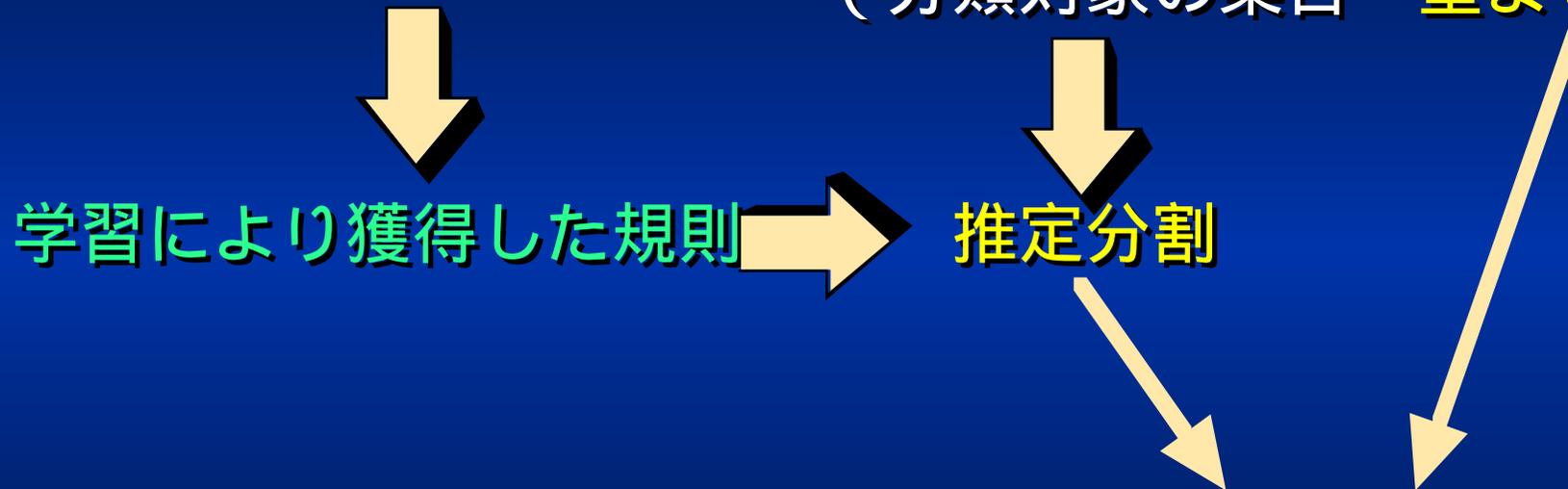
判別結果に
基づいた分割

学習結果の定量的評価

学習用の事例

テスト用の事例

(分類対象の集合 望ましい分割)



類似性を定量的に評価情報損失量 (RIL)

0 から 1 の範囲の値, 0 のとき完全に一致

定量的評価と定性的評価の対応

正しい分割と予測分割の二つの分割を可視化
情報損失量による定量的評価と共に示す

- 情報損失量が小さい方が定性的な評価でも優れる
- 正しい分割との視覚的一致が確認できる限界は 0.4 程度

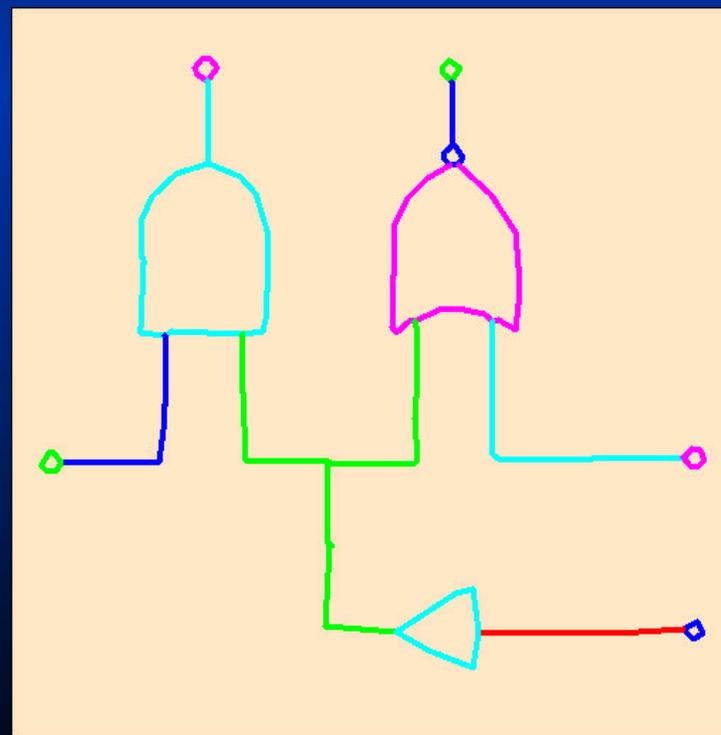
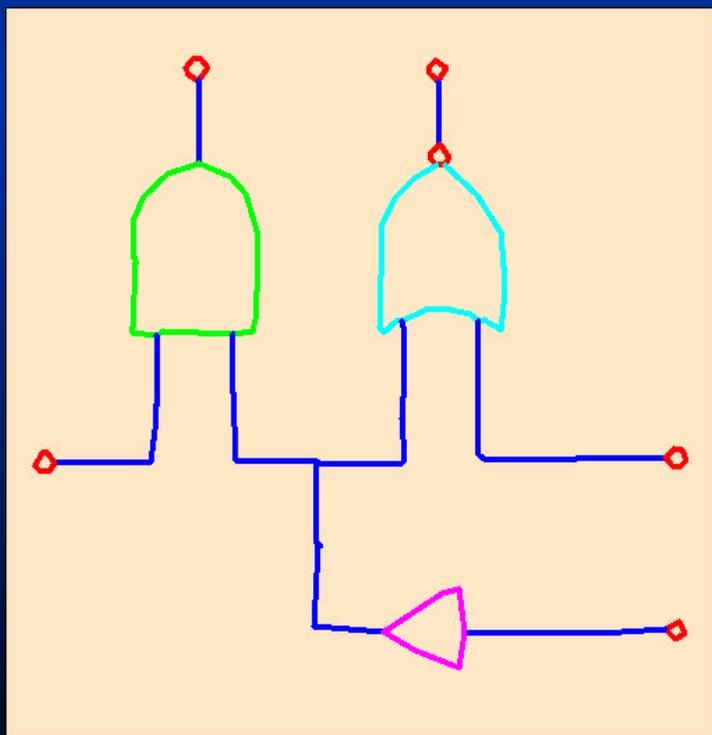
ベクトル・データの結果

論理回路・近傍の最も優れた結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.113



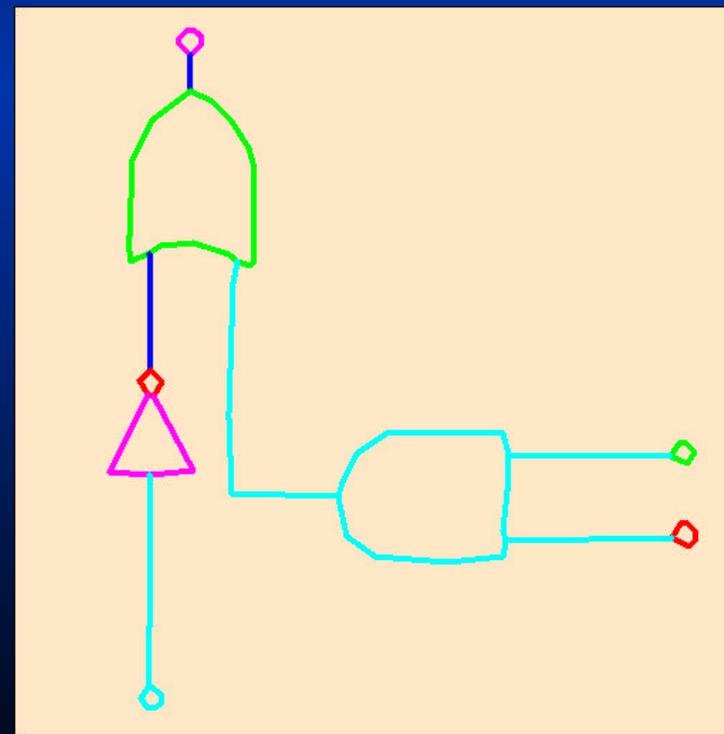
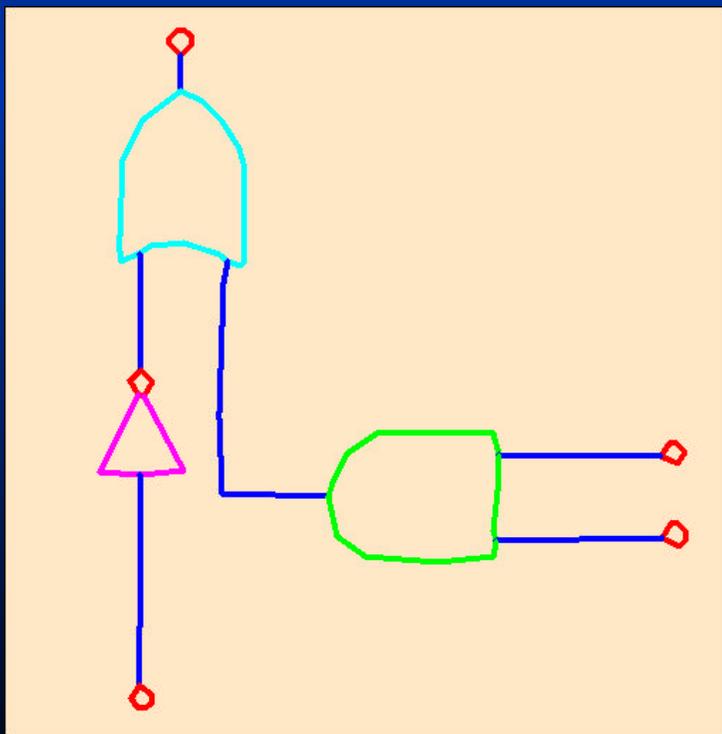
ベクトル・データの結果

論理回路・近傍の 10 番目に優れた結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.282



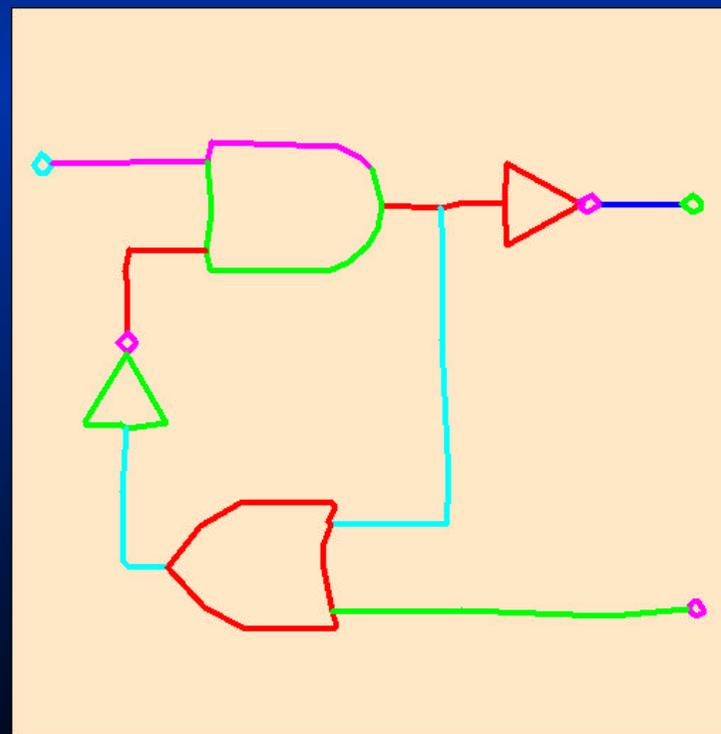
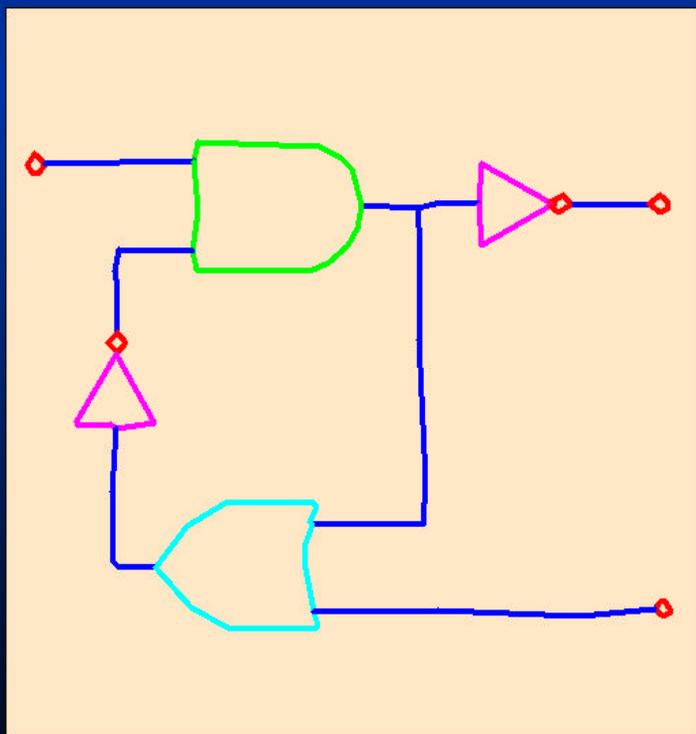
ベクトル・データの結果

論理回路・近傍の 20 番目に優れた結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.374



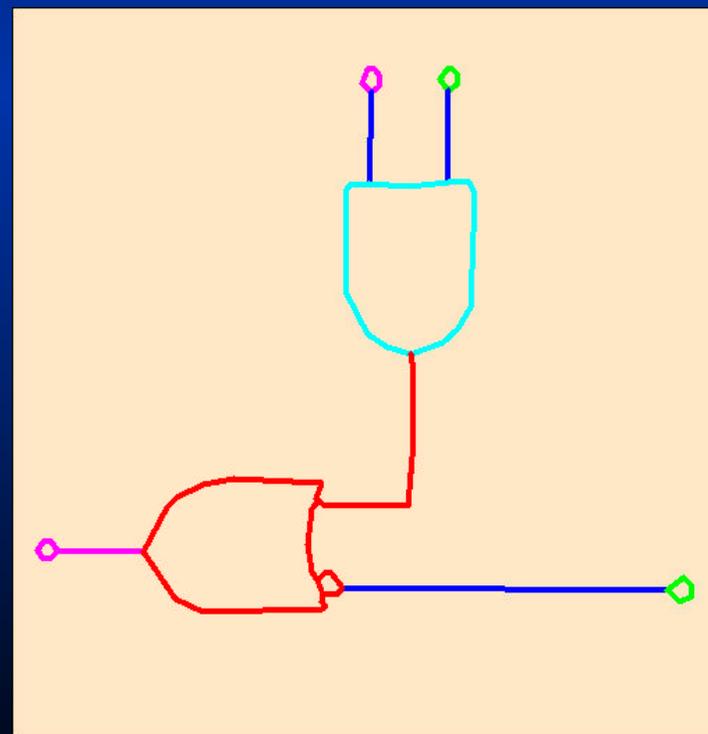
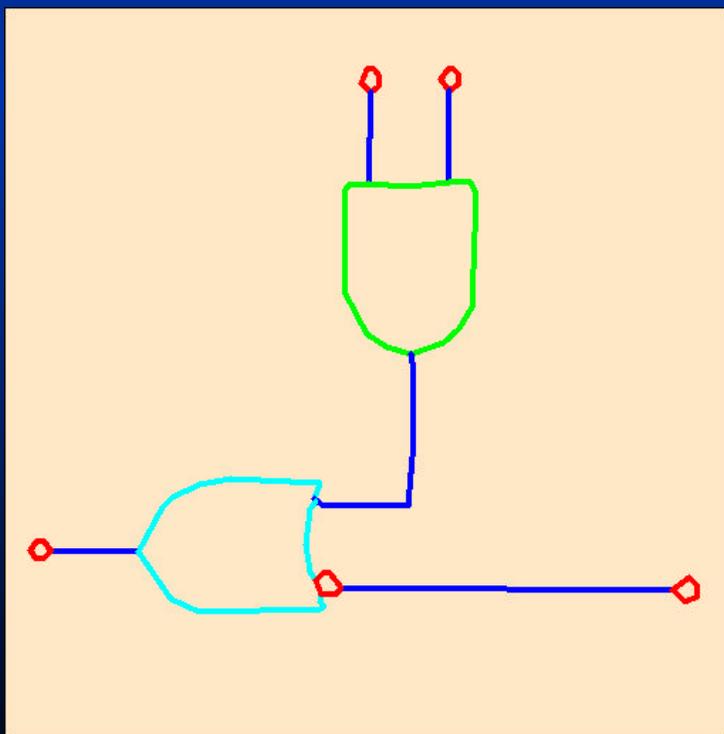
ベクトル・データの結果

論理回路・近傍の RIL が約 0.4 の結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.410



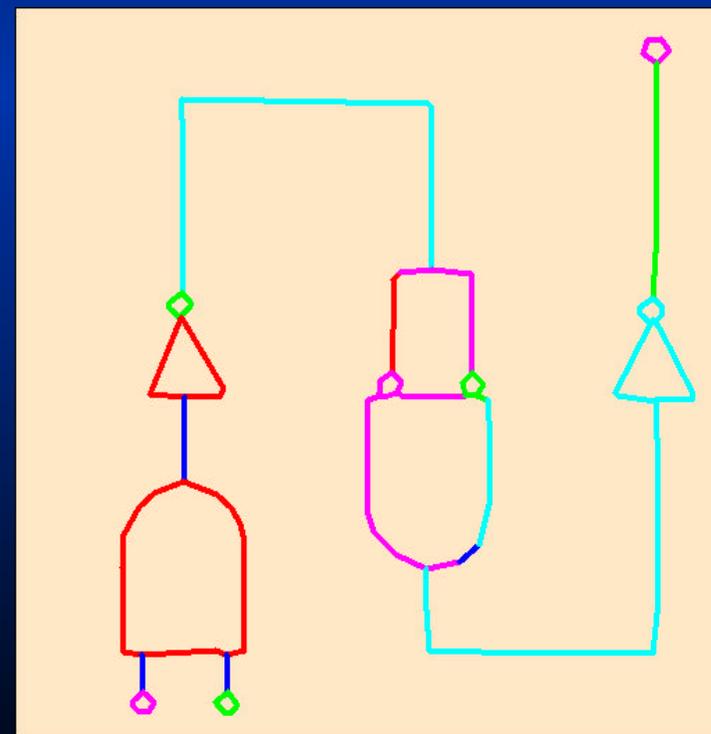
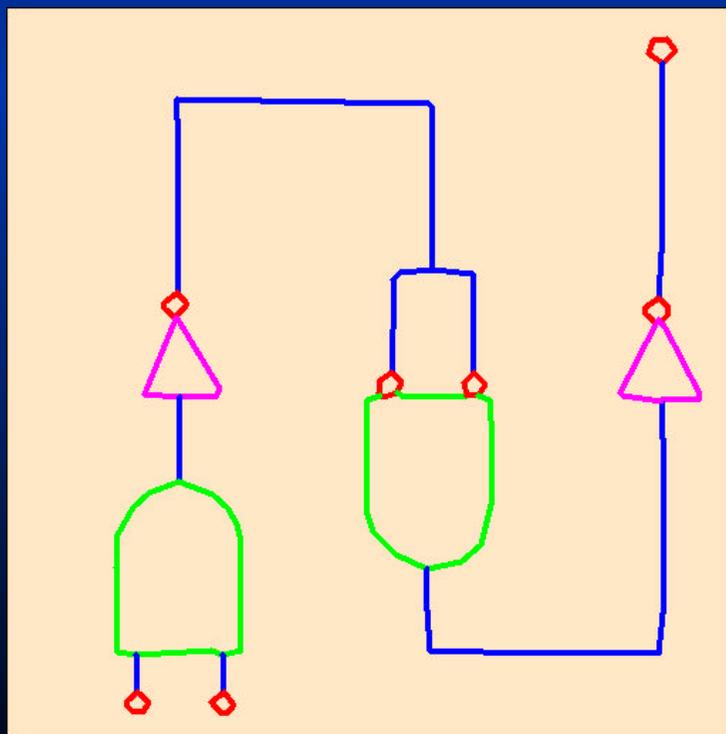
ベクトル・データの結果

論理回路・近傍の40番目に優れた結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.464



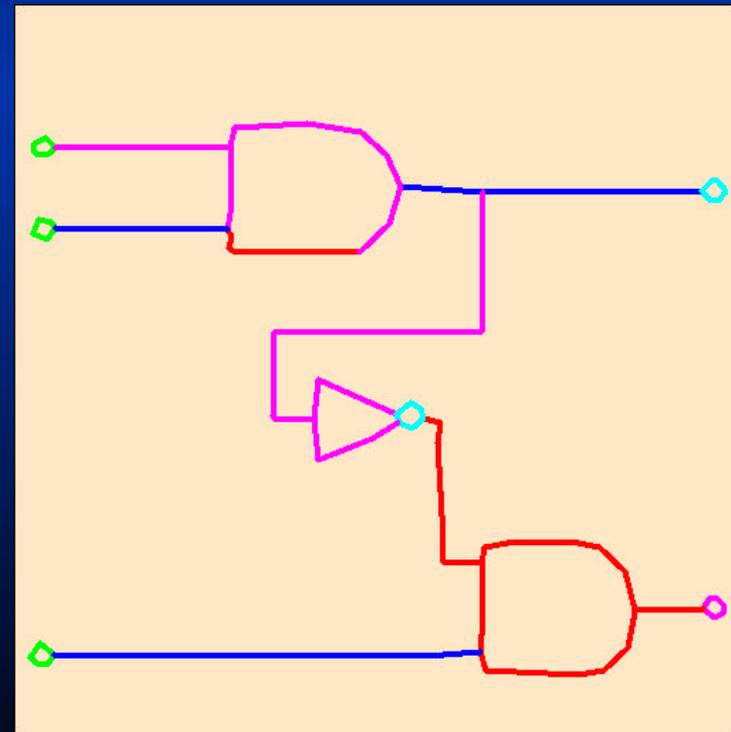
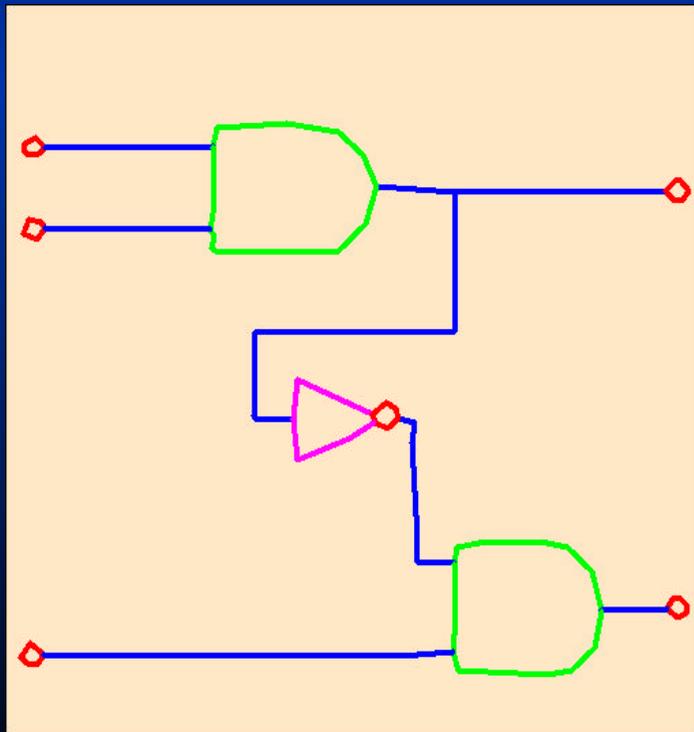
ベクトル・データの結果

論理回路・近傍の平均的な結果
(最小距離法を使用)

正しい分割

推定分割

RIL = 0.482



一般の例からの学習と クラスタリングに対する例からの学習

前提：実際の応用分野では、分類対象の対を完全に正しく判別することは困難

本研究の学習と一般の学習は本質的に同じ問題か？

分類対象の対に対する判別結果の改善



獲得できる分割の質の改善

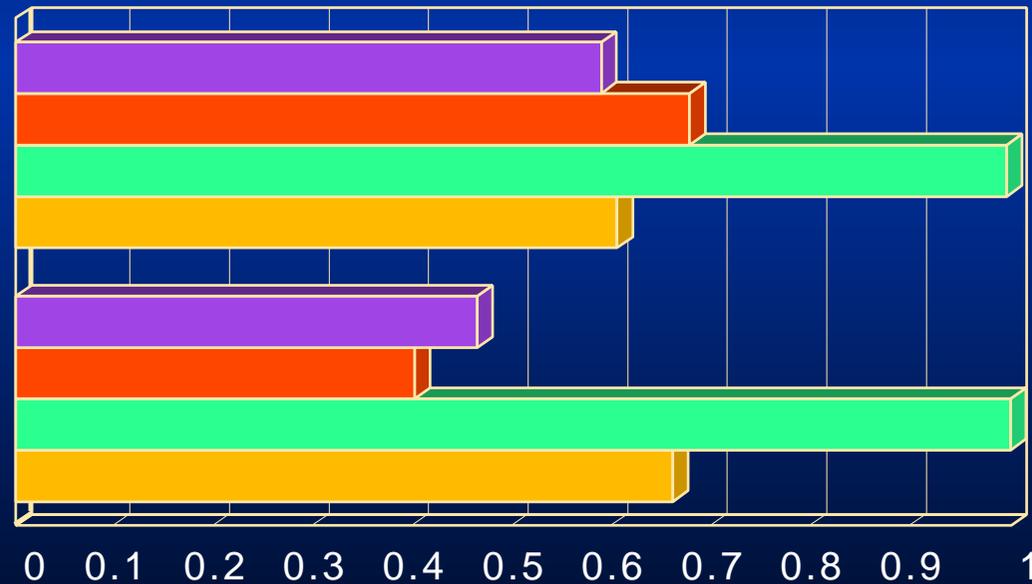
判別結果がよくても
望ましい分割が獲得できない事例を示す

データ「幾何図形」に対する 情報損失量

分割前：分類対象の対に対する判別結果
情報損失量により評価

分類対象の対の
選択基準

ランダム
近傍



分割前 最小距離法 最大距離法 群平均法

情報損失量の結果に対する考察

- 分割前 データ「近傍」がよい
- 
- 最小距離法 データ「近傍」がよい
 - 最大距離法 どちらのデータもだめ
 - 群平均法 データ「ランダム」がよい

分類対象の対の判別の他に
対の選択基準やクラスタリングの手法も重要

一般の例からの学習と
本研究の学習は本質的に異なる問題

まとめ

- 実験内容
 - 「クラスタリングに対する例からの学習」の学習法の提案と、その結果の定量的な評価
- 実験により示したこと
 - 定量的な評価と定性的評価との対応
 - 従来例からの学習と本研究の学習の相違
 - (既存のクラスタ分析手法を本学習に利用する場合の問題点)
- 今後の予定
 - 本学習に適したクラスタリング方法の開発

宣伝

WWW によるプログラムとデータの公開

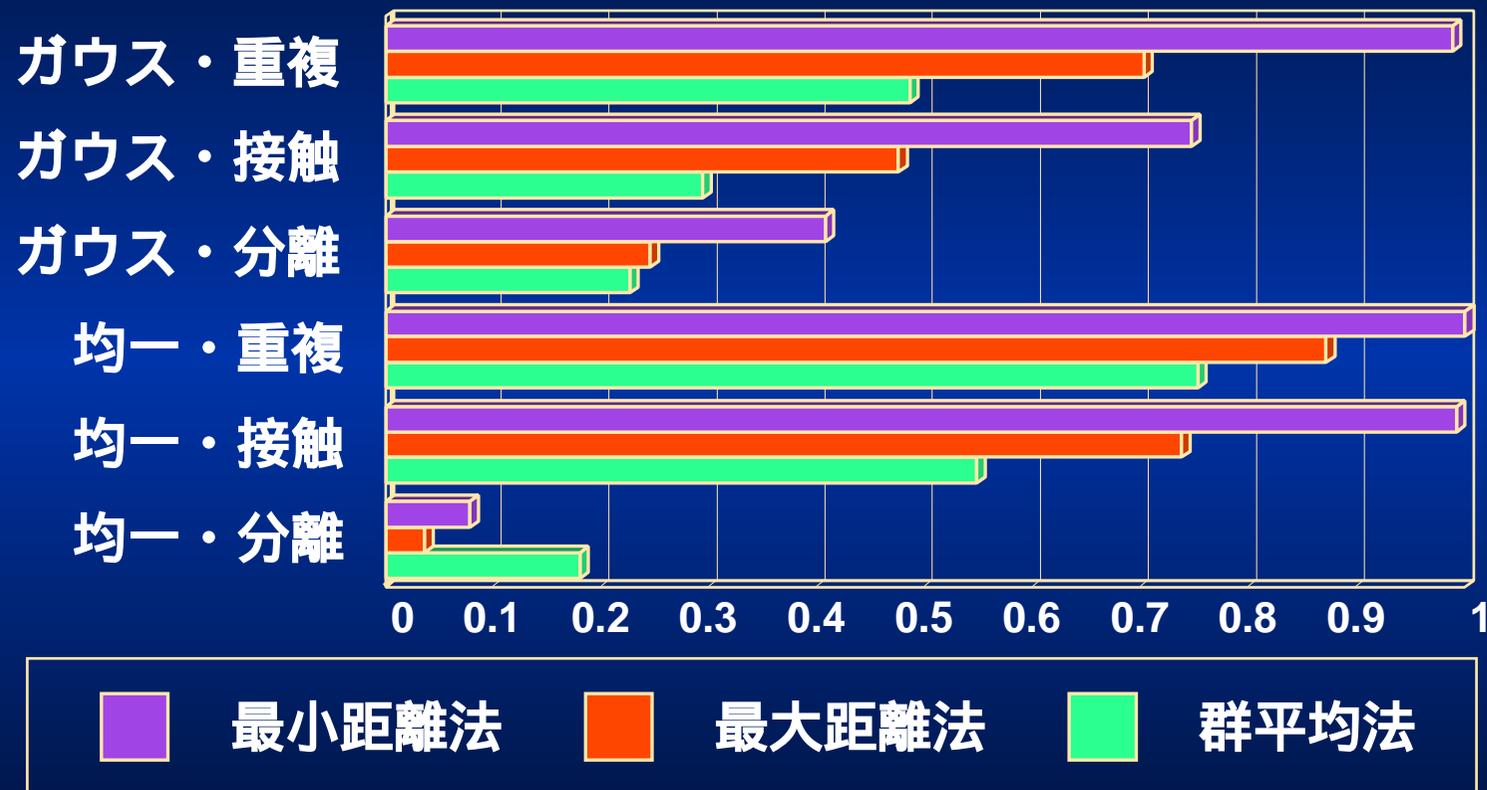
次の WWW のページに
本研究で用いたプログラムとデータがあります

[http://www.etl.go.jp:8080/etl/suiron
/ kamisima/J-HOME/program.html](http://www.etl.go.jp:8080/etl/suiron/kamisima/J-HOME/program.html)

SunOS 4.1.3 と HP-UX 9.0.1 で動作を確認しました

ご意見・ご感想をお聞かせ願えれば幸いです

ドット・パターンの実験



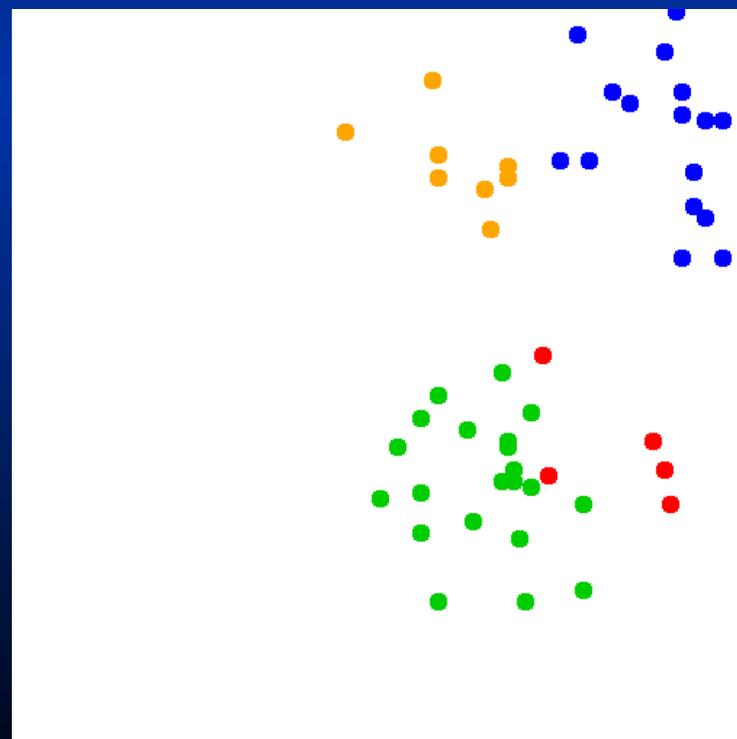
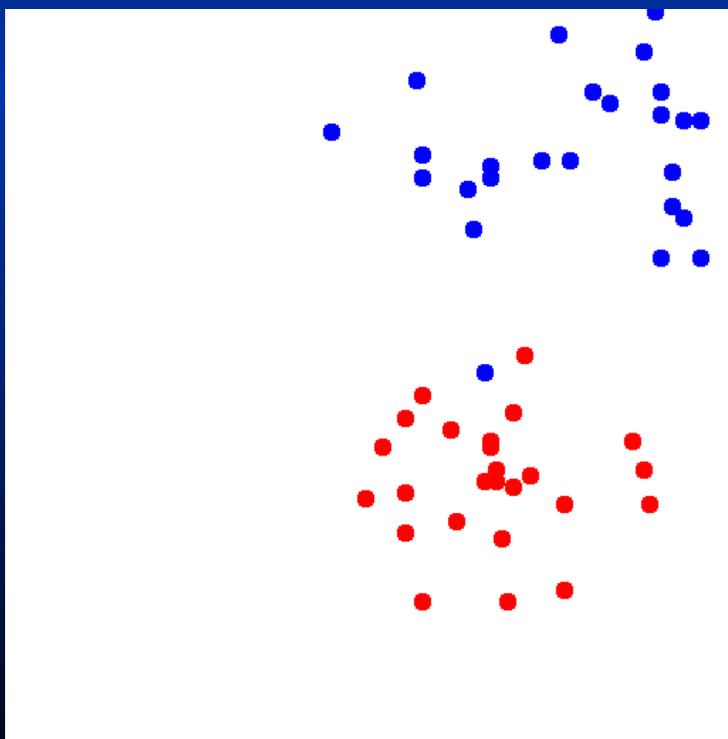
ドット・パターンの実験結果

ガウス・重複 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.703



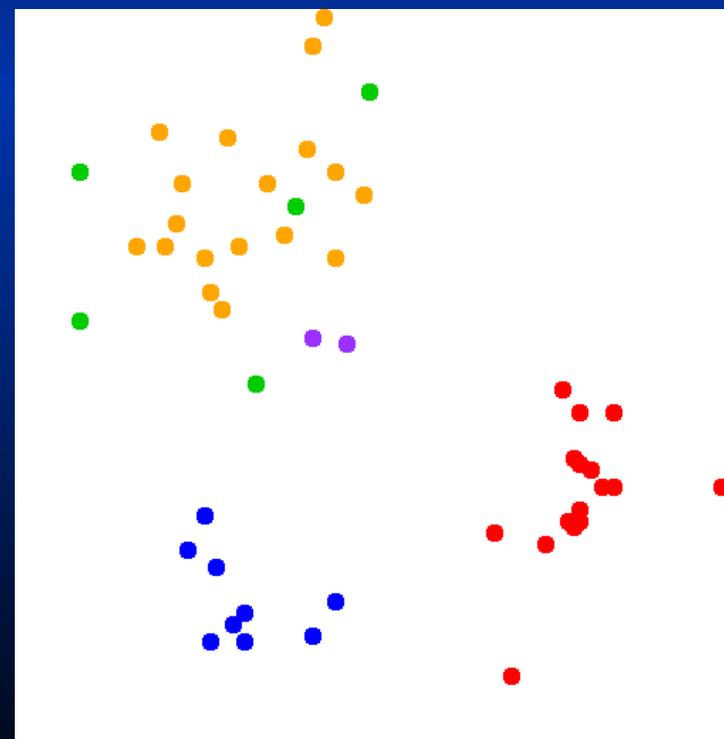
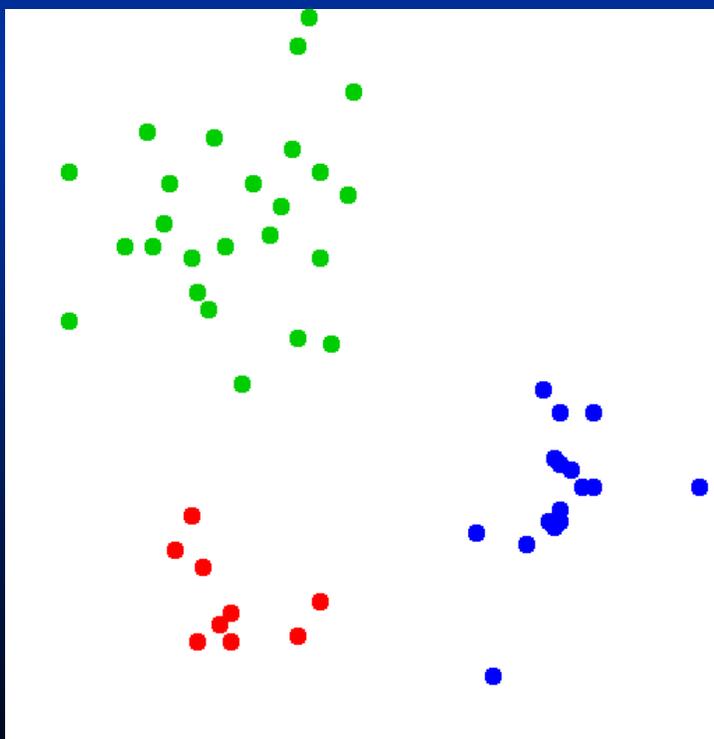
ドット・パターンの実験結果

ガウス・接触 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.474



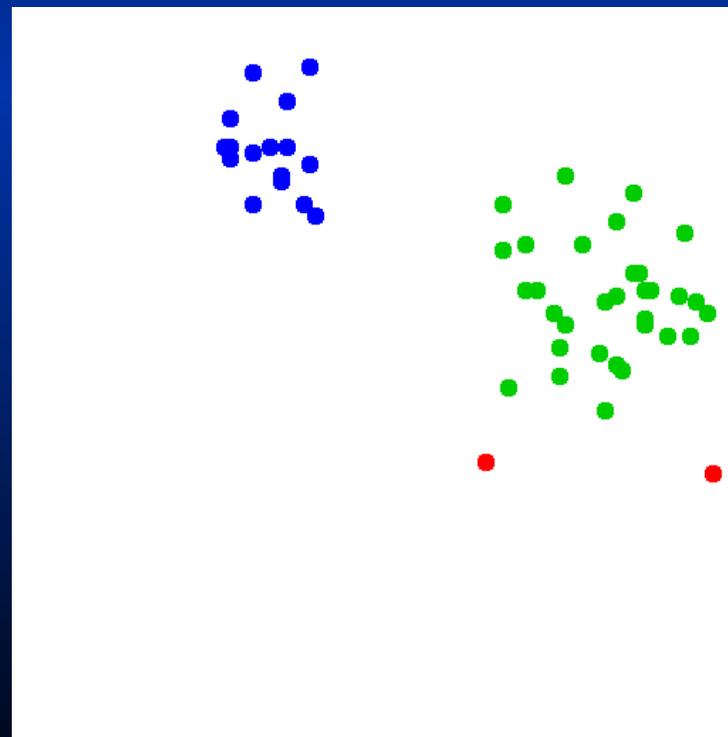
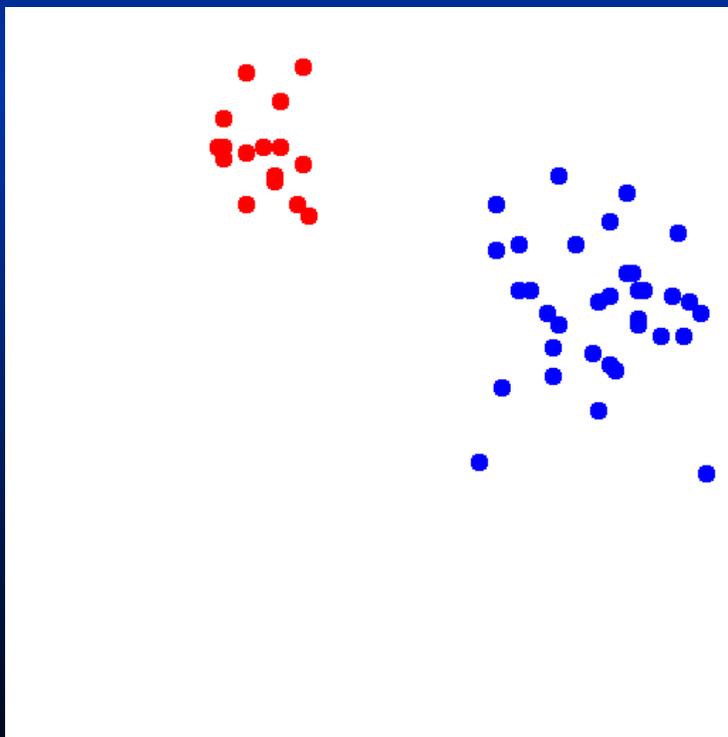
ドット・パターンの実験結果

ガウス・分離 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.243



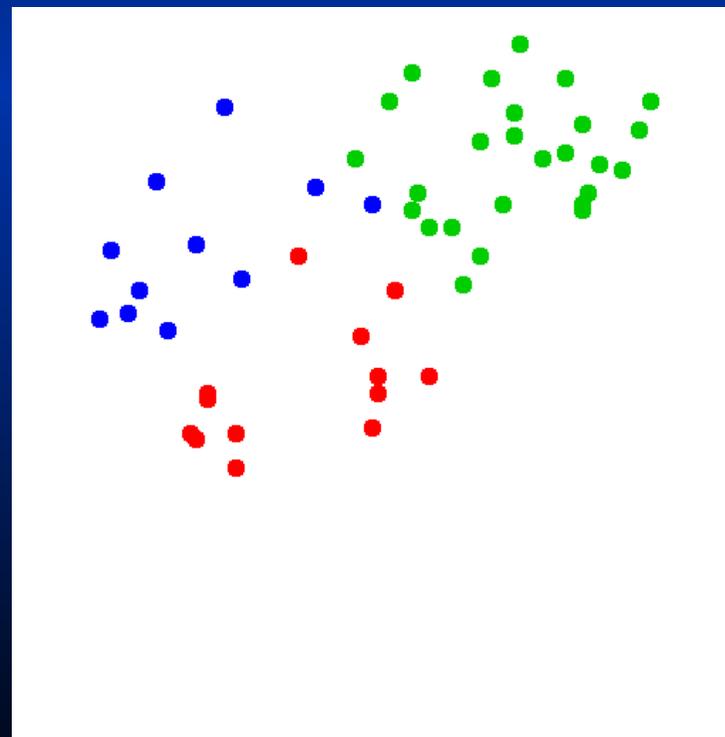
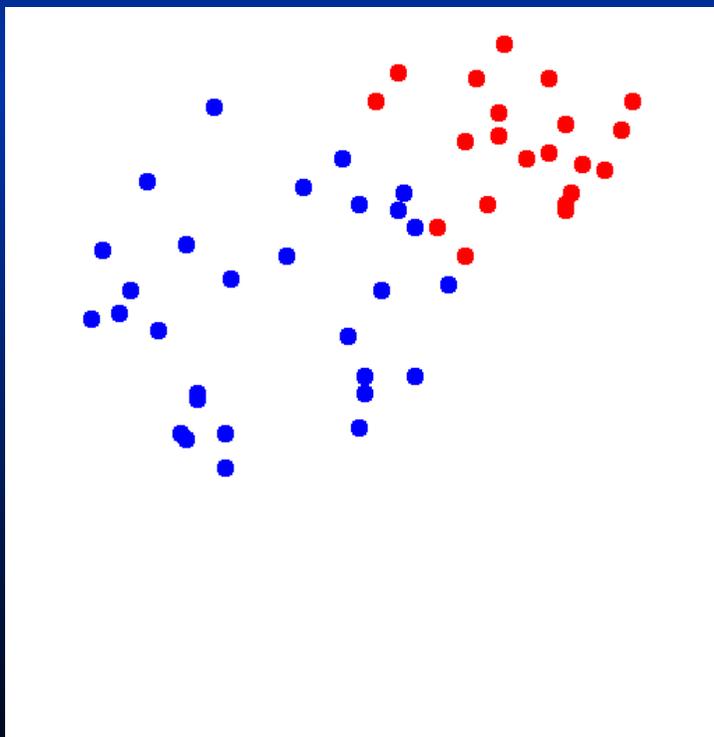
ドット・パターンの実験結果

均一・重複 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.871



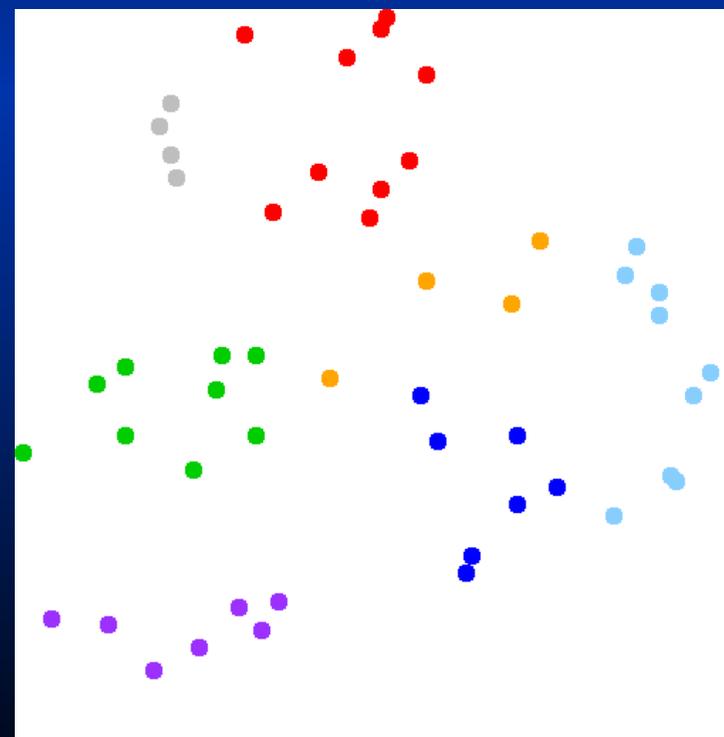
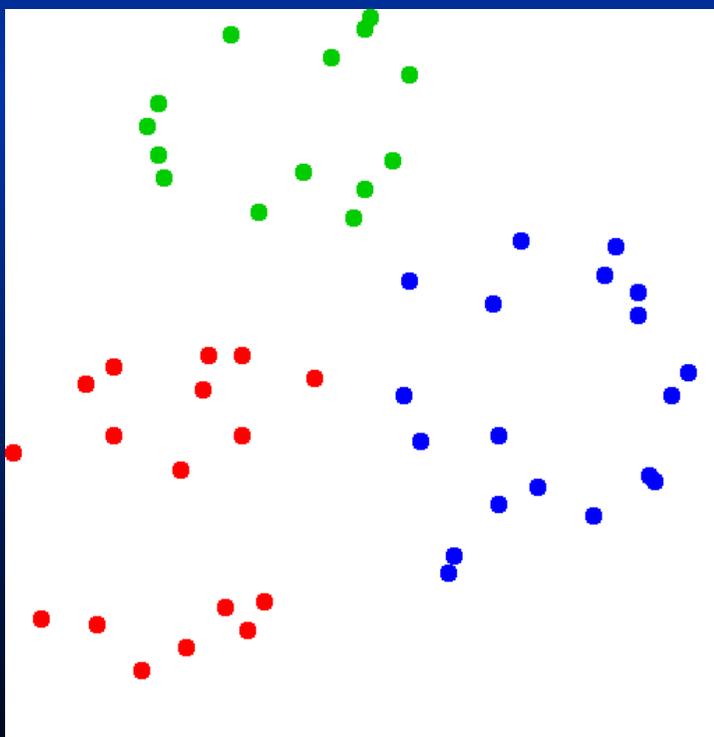
ドット・パターンの実験結果

均一・接触 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.735



ドット・パターンの実験結果

均一・分離 (最大距離法を使用)

正しい分割

推定分割

RIL = 0.043

