

研究タイトル：

大規模データのノンパラメトリック統計解析



氏名：	山本 けい子/YAMAMOTO Keiko	E-mail：	kei.yama@hakodate-ct.ac.jp
職名：	准教授	学位：	博士(工学)
所属学会・協会：	日本統計学会		
キーワード：	ノンパラメトリック, 確率密度関数, カーネル法, データマイニング		
技術相談 提供可能技術：	・大規模データの統計解析手法/データマイニング手法		

研究内容： 大規模データに対するノンパラメトリック統計手法の提案

大規模データの収集環境の整備にともない、収集したデータを活用するための解析技術が必要になってきている。データマイニングや機械学習など盛んに研究が進められており、本研究では、統計学の立場で解析する技術の開発を目指している。

データのもつ確率構造を柔軟に表現する統計的手法の1つとして、あらかじめ分布を仮定しないノンパラメトリック法による確率密度関数の推定法がある。

高次元大規模データのような複雑な現象に対して、確率構造を表現するための統計モデル構築には、ノンパラメトリックな手法が適していると考えられるが、高次元データに対しては、計算負荷などの点からあまり実用化されておらず、理論的な性質は未整備である。

そこで、本研究では、高次元大規模データを扱うためのノンパラメトリック統計解析手法を提案し、その理論的性質を明らかにする。

主要な研究テーマは、「ノンパラメトリック統計手法による非線形判別器の構築」で、手書き数字認識問題を例に、密度関数をベースにした非線形判別器を提案する。ノイズのあるデータに対して、本手法が頑健である結果が得られており(図 1)、さまざまなデータを対象に実用化に向けて研究を進めている。

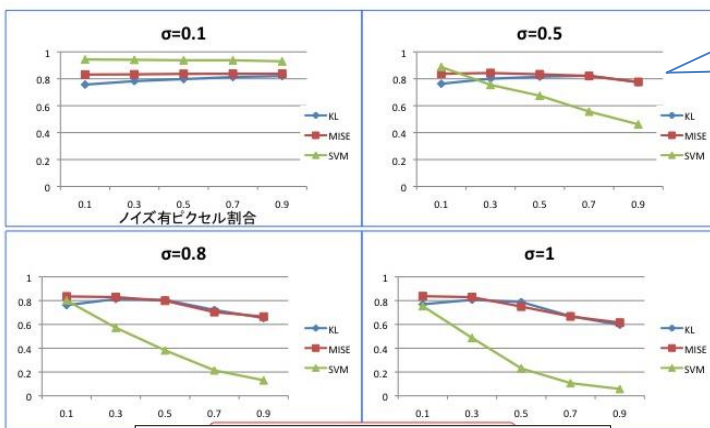


図 1. ノイズありデータに対する判別精度

提案手法(評価方法 MISE または KL) はサポートベクターマシン(SVM)に比べ精度の低下がゆるやか

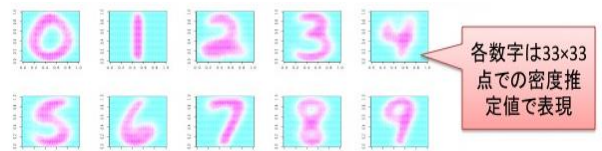


図 2. トレーニングデータを用いた正解数字密度分布

提供可能な設備・機器：

名称・型番(メーカー)