

## クラスタ数推定に用いる最適な情報量基準の探求

出席番号： 32 番      報告者： 萩原 涼介      指導教員： 藤田 一寿      提出日： 2017 年 6 月 15 日

### 1. はじめに

クラスタリングとはデータを教師なし学習により任意の数のクラスタに分ける手法である。クラスタリングはデータ解析，データマイニング，パターン認識など様々な分野で用いられる。多くのクラスタリング手法では，予めクラスタ数を指定しクラスタリングを行う。しかし，データに対し最適なクラスタ数を指定しなければ，最適なクラスタリング結果を得ることはできない。その為，クラスタ数を推定することは重要な課題となっている。

既存のクラスタ数推定手法の多くは，情報量基準に基づきクラスタ数の推定を行っている。情報量基準とは簡単に言えば確率分布とデータの分布の当てはまり具合を表す。その情報量基準は多くの研究者により様々なものが提案されている。しかし，どの情報量基準がどのようなデータに対し有効かは分かっていない。そこで本研究では，クラスタ数推定に用いる情報量基準として最適なものを数値実験を通し明らかにする。

### 2. 実験の経過

### 3. 実験の手法

### 4. 結果

### 5. おわりに

### 参考文献