

多様体の学習における話(考察編)

芝浦工業大学 数理科学研究会

bv18079 山岡 史弥

平成 30 年 11 月 2 日

1 研究背景

前期に機械概論という授業をとった。この授業では微分幾何学における多様体学習や多様体学習の実装を扱った。この授業を通じて多様体学習に興味を持ったので、研究のテーマとして取り上げた。

今まで私たちは高次元空間に埋め込まれた多様体を研究してきた。特に線形変換を用いた主成分分析によって低次元データをよく説明しようと試みることで、非線形データの低次元で表現することが可能となる。次元削減においても同じ事がいえるが主成分分析のうまい座標変換を施して高次元から低次元データを説明したい。

2 Isomap について

Isomap とは、非線形次元削減のひとつであり、 k 近傍グラフを用いて、多様体上測地線距離を求め、多次元尺度構成法を用いる。近似的に Euclid 的な低次元空間に射影するものである。

3 非線形の難しさ

非線形次元削減という手法は複雑で、かつ結果の解釈も難しくなりがちである。一方、線形次元削減の手法は比較的簡単で、かつ結果の解釈も簡単である。例えば、主成分分析の結果の解釈などがあげられる。

4 Isomap の難しさ

k 近傍グラフの k の値を考える。これは、 k が小さすぎると全体が正しく連結されず、 k が大きすぎると多様体の構造が正しく反映されないからである。

5 Isomap でうまくいくための近似の条件

多次元尺度構成法の精度は距離行列の固有値が正であるとき最大となる。ここで、距離行列とは、与えられた点

データ間の距離を並べて行列にしたものである。例えば、距離関数が Euclid 距離であるとき、距離行列の固有値はすべて正となる。

そこで、Isomap では、負の値を無視して、固有値を大きいほうから 2, 3 個用いる。

距離がユークリッドでないときにおいては一般に正の固有値が現れる Isomap において負の固有値は無視して固有値の大きい順にいくつか (2, 3 次元) を採用する測地線距離のユークリッド性が弱いと近似が悪くなる

$$1. D_{ij} = (d_{ij})^2$$

$$2. K = -\frac{1}{2} \cdot HDH$$

$$3. H_{ij} = \delta_{ij} - \frac{1}{n} \cdot I_i$$

のようになる

6 今後の課題

今回、多様体学習について学ぶことができた。しかし、多様体については深く学ぶことができなかった。そこで、今後は多様体について、勉強していきたい。

参考文献

- [1] 改訂新版:代数曲線形幾何学, 東京出版.
- [2] <https://www.slideshare.net/kohta/risomap>, 2011 年 5 月 28 日