データサイエンス

keu word

- 統計学
- 統計解析
- 非負値行列因子分解
- テキスト解析

佐藤 健· Kenichi Satoh

データサイエンス学部 教授

【プロフィール】

- ·広島大学大学院理学研究科 (博士課程前期)数学専攻 修了
- ·広島大学原爆放射線医科学 研究所 計量生物研究分野 准教授
- ・滋賀大学データサイエンス学 部 教授 2021 年 4 月~継

【主な社会的活動】

日本統計学会, 応用統計学会

【主な受賞歴】

・2010年度, 2015年度および 2022 年度応用統計学会 学 会賞(優秀論文賞)受賞

【連絡先】

kenichi-satoh@biwako.shigau.ac.jp



【代表的な研究テーマ】

□ 統計学

課題解決に役立つシーズの説明

【研究紹介】

非負値行列因子分解(NMF)は Lee & Seung (Nature, 1999)によって顔認識の手法として 提案されたパーツベースの手法である。データの構成要素と割合を求めて解釈を容易に する. ここではケーキを例に紹介する.

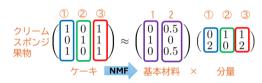
1. ケーキの特徴をざっくりと数値化してみる





- ・クリーム多い スポンジ多い
- 果物あり

2. ケーキデータにNMFを適用する



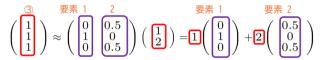
- # install.packages("remotes")
- # remotes::install github("ksatohds/nmfkc") Y <- cbind(c(1,0,1),c(0,1,0),c(1,1,1))
- library(nmfkc)
 res <- nmfkc(Y,Q=2,epsilon=1e-6)</pre>
- res\$X; res\$B

R言語によるNMFの実装

ーキデータをNMFによって基本材料と分量に分けた.基本材料としては材料1 {スポンジ} , 材料2 {クリーム,果物} が 抽出された. したがって、3つのケーキはこの2つの材料を定数倍して加えて近似できる.



材料1を分量1,材料2を分量0で近似できる. つまり,ケーキ②は材料1だ +2は, けで構成されていると解釈できる.





講義資料

ゲーキ③は, 材料1を分量1,材料2を分量2で近似できる.また,分量の割合を考えれば,材料1が1/(1+2)=33%,材料2が2/(1+2)=67% と考えられるので,どちらかと言えば,材料2の寄与が高いと解釈できる.このように,NMFはデータから構成要素を抽出して,近似・分類する.

【著書紹介】

実況!R で学ぶ医療・製薬系データサイエンスセミナー

学術図書出版社: https://www.gakujutsu.co.jp/product/978-4-7806-1103-8/

企業・自治体へのメッセージ

- アカデミックな共同研究の経験が豊富です。
- ・統計ソフト「R」を使って色々な統計解析手法を使ったり、説明したりできます。