

疎行列の非負値行列因子分解のための 効率的な近似推定法

阿部興¹ ・ 島村徹平¹

2024 年 10 月

¹東京科学大学 総合研究院 難治疾患研究所

動機

次元の呪い \longrightarrow 次元の祝い²

- 高次元の積分は難しい
- △ 高次元のデータ分析はややこしい
- ◎ 見えるもの（変数）が多いことは嬉しい！ たとえそれが“0”だったとしても

疎（sparse）なデータ \neq 汚いデータ

²Gelman, A. (2004). “The blessing of dimensionality”
https://statmodeling.stat.columbia.edu/2004/10/27/the_blessing_of/

非負値行列因子分解（NMF）の紹介

- 得られたデータを低次元に射影して圧縮することでパターンを抽出
- 非負性の制約により解釈がしやすい

疎行列と NMF

- 観測のゼロ過剰や過分散をモデル化したケース³
- 分解で得られる行列を疎にしようとした議論⁴

³e.g. Abe, H., & Yadohisa, H. (2017). A non-negative matrix factorization model based on the zero-inflated Tweedie distribution. *Computational Statistics*, 32(2), 475-499.

Gouvert, O., Oberlin, T., & Févotte, C. (2020). Negative binomial matrix factorization. *IEEE Signal Processing Letters*, 27, 815-819.

⁴e.g. Hyunsoo, K. & Haesun, P. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, *Bioinformatics*, 23 (12), 1495–1502.

Kim, H., & Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, 23(12), 1495-1502.