

一般的に認知症や生活習慣病などの多因子疾患は遺伝的要因や社会科学的要因、生活習慣などが関連しあって生じる。実際に認知症に関しては、糖尿病や高血圧をはじめとする生活習慣病や身体的活動、喫煙等がリスク上昇に寄与することが示されており、発症に関連する SNPs や腸内細菌環境に関する研究も多く行われている。そのため、多因子疾患に対する予測モデルを作成する、あるいは疾患と関連する要因間の関係性を把握するためには、ゲノムデータから社会環境要因までを含めた多角的な情報をもつデータを解析する必要性がある。

ゲノムデータのように変数が多いデータを高次元データという。高次元データ解析の初期段階ではデータの可視化が必須であり、その一手法としてグラフィカルモデルが用いられる。グラフィカルモデルとは確率モデルをグラフ化して表したモデルであり、データを視覚化し変数間の関連構造を把握することや予測など多くの用途をもつ。理論面では、グラフィカルモデルは構造方程式のように想定するモデルをもとに変数間の関連度合いを調べる多変量解析の一手法として用いられてきた。一方で、近年の計算技術の発展とともにデータからグラフ構造を決定する方法の研究が進められている。そこではベイジアンネットワークに代表されるように、尤度関数に基づくベイジアン理論を基礎とした機械学習手法として位置づけられるようになっている。

本抄読会では、解析対象とするデータについて簡単に説明し、グラフィカルモデル理論の基本的事項と応用例、また欠測値の補完方法について述べる。

参考文献

1. Scutari M, Denis JB. *Bayesian networks with Examples in R*. Chapman & Hall: CRC press, 2014.
2. Rodin A, Mosley TH Jr, Clark AG, Sing CF, Boerwinkle E. Mining genetic epidemiology data with Bayesian networks application to APOE gene variation and plasma lipid levels. *Journal of Computational Biology* 2005; **12**:1-11.
3. Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; **28**:112–118.