

# 主成分分析

## 概要

- 英語ではPCA (Principal Component Analysis)
- 教師なし学習
- 次元削減を行う
- データの分散が大きいものを主成分とする
- 例：乳がん検査データについて32次元データを2次元に次元圧縮

## 寄与率

- 第k主成分の分散の全分散に対する割合(第k主成分が持つ情報量の割合)
  - 第k主成分がデータ全体のどれだけの分散をカバーしているかを示す

## 累積寄与率

- 第1~k主成分の分散の総和が全分散に占める割合
  - 第1~第k主成分までで全体のどれだけの分散をカバーしているかを示す

## 乳がん検査データのハンスオンでの例

- 第1主成分の寄与率: 0.43315126
- 第2主成分の寄与率: 0.19586506
- この場合、第2主成分までの累積寄与率は0.62901632となり、2次元に圧縮しても約6割の情報量があることが分かる。