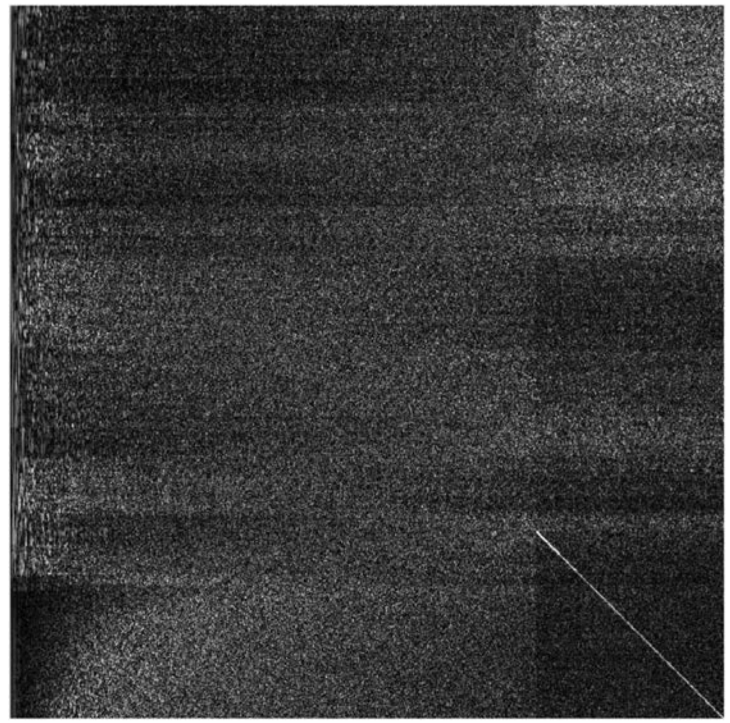
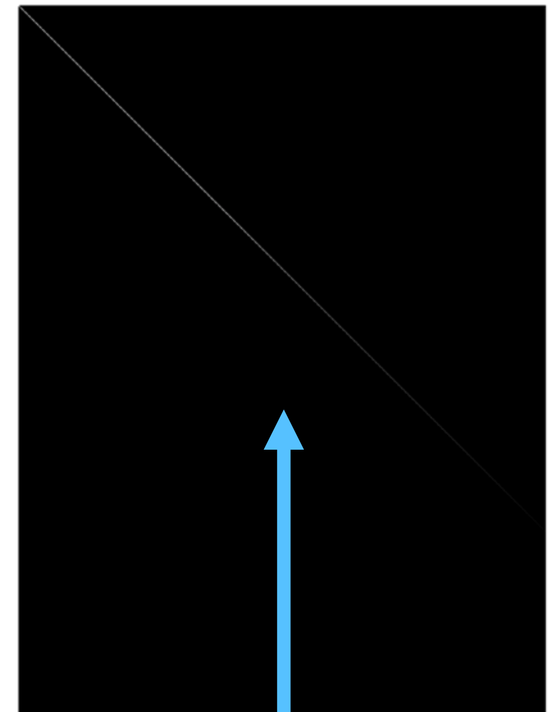
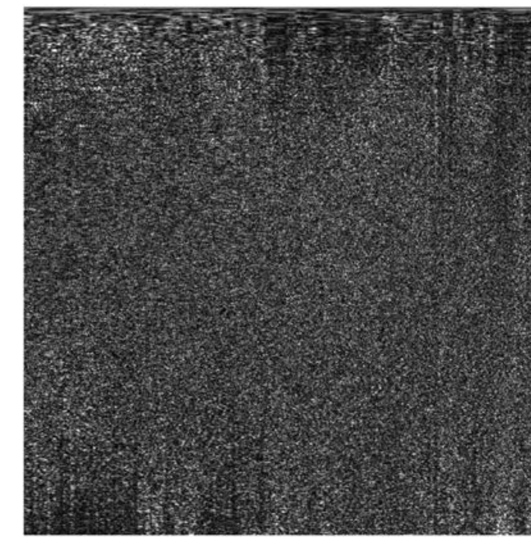
数理手法VIII

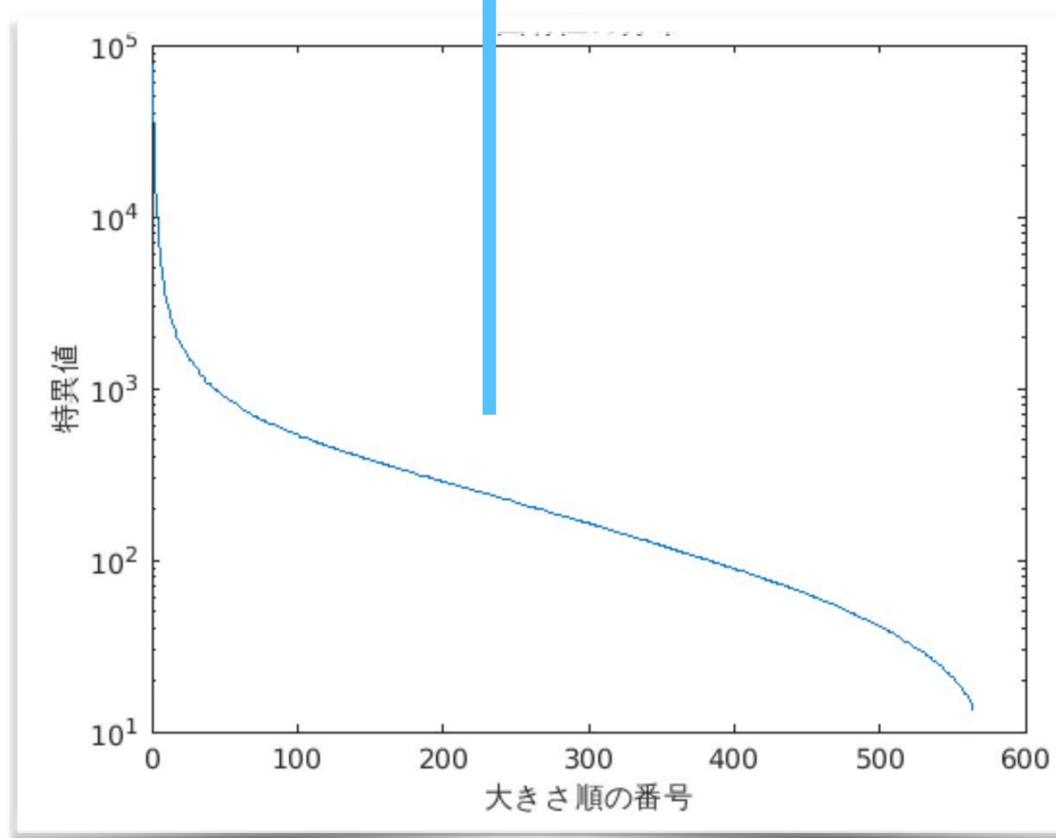
数理・情報教育研究センター 数学基礎教育部門
工学系研究科システム創成学専攻

島田 尚

画像の特異値分解

A  $=$ U  Σ  $V^\top$ 

（要素の絶対値で表示）

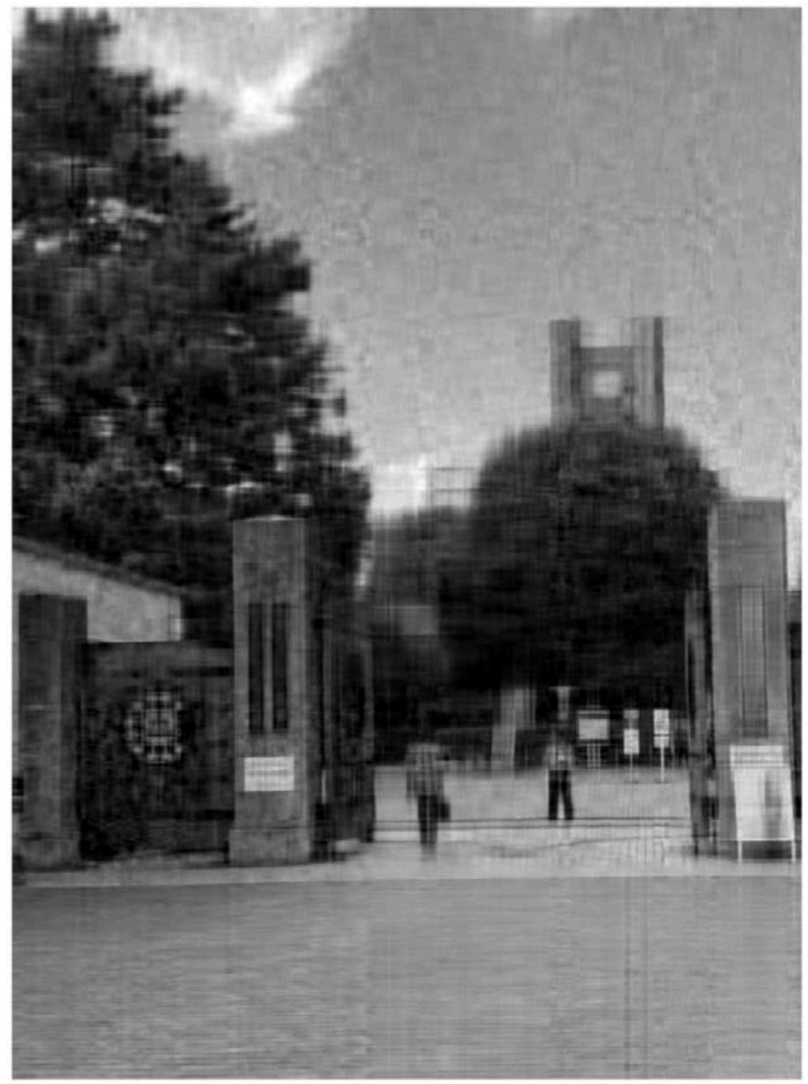


フルランクの特異値をとらなくても、
結構良い再構成ができるのでは？
↓
画像の圧縮
（特異値分解の低ランク近似）

特異値分解の低ランク近似による画像の圧縮

100次元

30次元

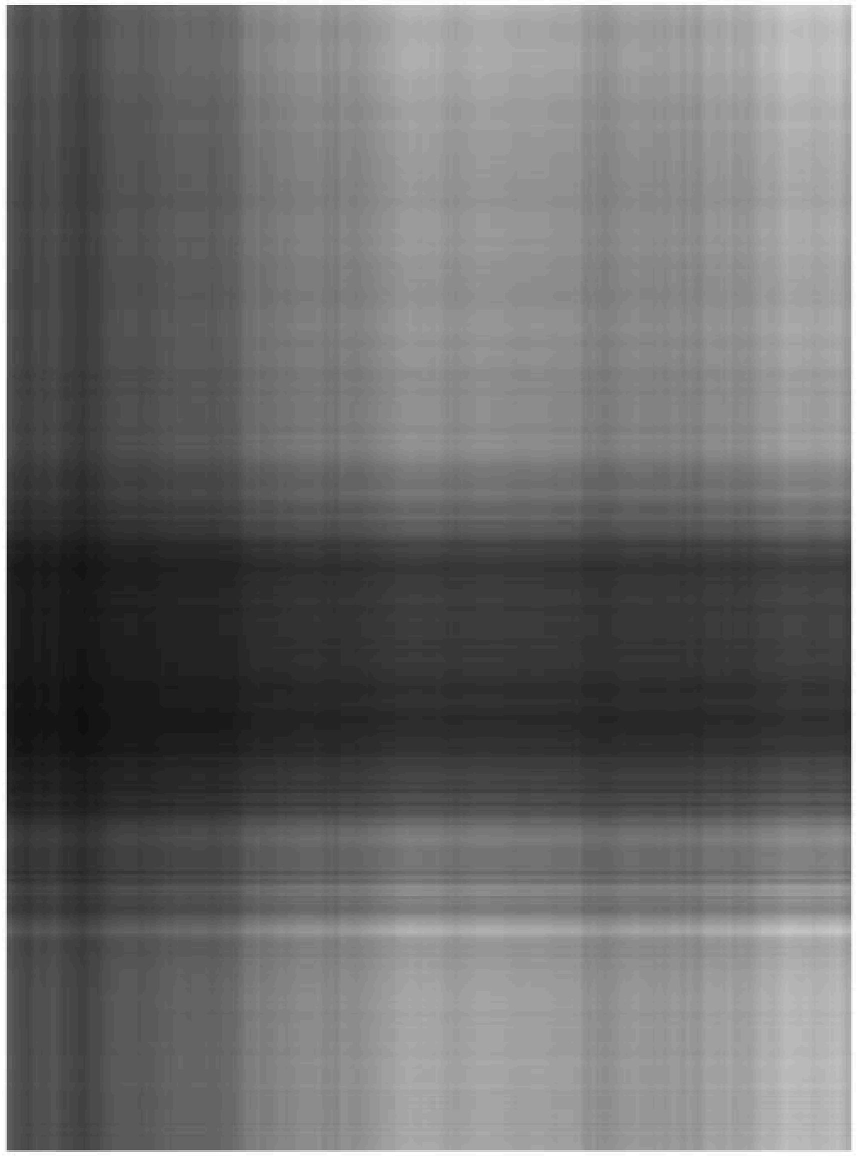


10次元

特異値分解の低ランク近似による画像の圧縮

（1次元）

$\vec{u}_1$



=

$\vec{v}_1^{\top}$

二項積（テンソル積、dyadic）

$$\left(A = U\Sigma V^{\top} = \sum_i \sigma_i \left(\vec{u}_i \vec{v}_i^{\top} \right) \right)$$

最小2乗問題と正規方程式

$\vec{x}_*$ が$\| A\vec{x}_* - \vec{b} \|_2^2$ を最小にする → 任意の方向の微小変位 $\vec{\epsilon}$ について

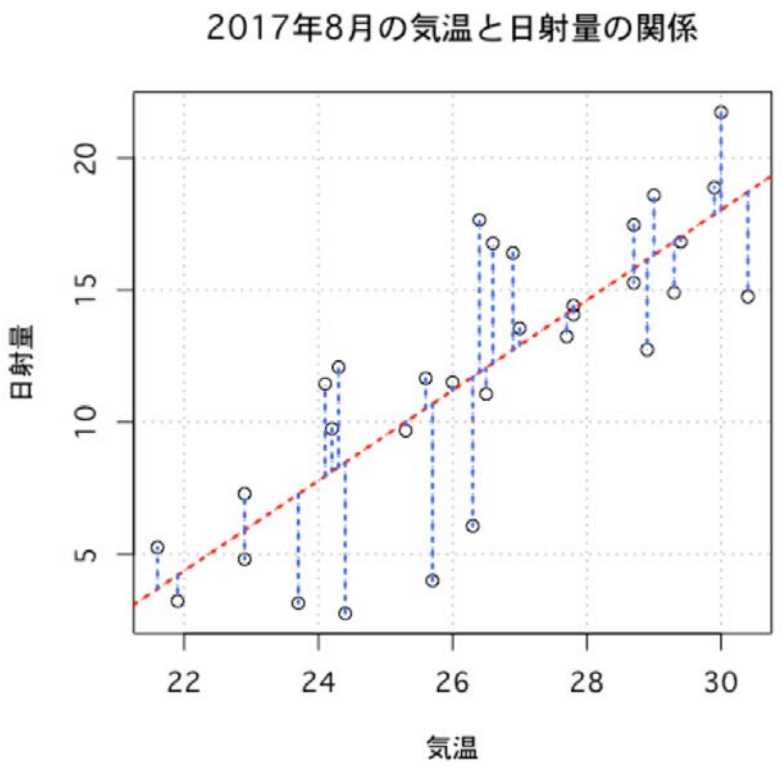
$$\| A\left(\vec{x}_* + \vec{\epsilon}\right) - \vec{b} \|_2^2 - \| A\vec{x}_* - \vec{b} \|_2^2$$

$$= \vec{\epsilon}^\top (A^\top A)\vec{\epsilon} + 2\vec{\epsilon}^\top \left[A^\top \left(A\vec{x}_* - \vec{b}\right)\right] \geq 0$$

$$\downarrow$$

<u>正規方程式：$A^\top A\vec{x} = A^\top \vec{b}$</u>

最小2乗法と
関数フィッティング（線形回帰分析）



y（目的変数）$\sim \alpha + \beta x$（説明変数）

✓最小二乗法:
モデルとデータとの差（残差）の二乗和 $\sum_{i=1}^{n}[y_i - (\alpha + \beta x_i)]^2$
が最も小さくなるようにαとβを決める

（多項式の場合）$y \sim \alpha + \sum_{k=1}^{n} \beta_k (x_i)^k$

関数をモデルと思うと、推測統計の視点からは何をしていることになるか？

↓

最尤法

尤度（モデルの尤もらしさ）と最尤法

<u>例</u>：表の出る確率を知らないコイン n 枚を投げたら、表の出た枚数が m 枚だった。
このコインの表が出る確率 p はいくつだと推定するのが良いか？

<u>尤度の考え方</u>:
表の出る真の確率を p と仮定すると、その場合に表が m 枚出る確率は二項分布より

$$\left(\frac{n!}{m!\,(n-m)!}\right)p^m(1-p)^{n-m}$$

表の出る確率を p_1 としたときと p_2 としたときの<u>実現確率の比</u>は

$$L(p_1)/L(p_2), (L(p)=p^m(1-p)^{n-m})$$

つまり、<u>この**L(p)**が最大になる **p** が仮説として一番尤もらしい</u>。
このような量 L(p) を<u>尤度</u>、尤度の最大値を<u>最大尤度</u>、
このようにパラメターp を決めることを<u>最尤推定（最尤法）</u>という

最尤法

例：表の出る確率を知らないコイン n 枚を投げたら、表の出た枚数が m 枚だった。
このコインの表が出る確率 p はいくつだと推定するのが良いか？

尤度: $L(p) = p^m(1-p)^{n-m}$ *を最大化*

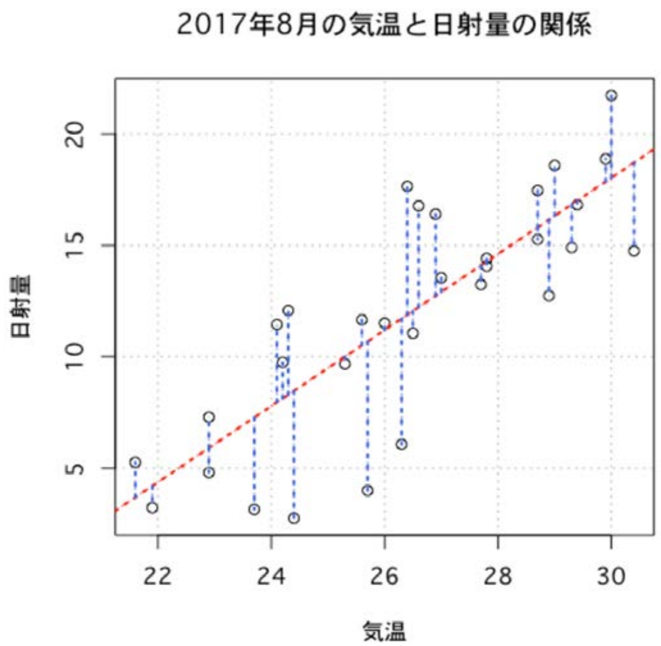
↓

対数尤度: $\log L(p) = m\log p + (n-m)\log(1-p)$ *を最大化*

$$\frac{d}{dp}\log L(p) = 0 \to p = \frac{m}{n}$$

“常識”や標本平均の推定の方法と一致

最尤法と線形回帰



$$y_i = \alpha + \beta x_i + \eta_i, \eta_i \sim N(0, \sigma^2)$$

1. 日射量の観測値 ＝「法則 ＋ ノイズ」と考え
2. ノイズに正規分布を仮定する

Photo by Deutsche Bundesbank, Frankfurt am Main, Germany, from Wikipedia
https://commons.wikimedia.org/wiki/File:10_DM_Serie4_Vorderseite.jpg
Public domain

法則（線形モデル）: $\alpha + \beta x_i$ によって
今持っているデータ $\{y_i\}$ が実現する確率（尤度）は

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\eta_i^2}{2\sigma^2}\right\} \propto \prod_{i=1}^{n} \exp\left\{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right\}$$

最尤法と線形回帰

今持っているデータ $\{y_i\}$ に対する対数尤度は

$$\log L = \log\left(\prod_{i=1}^{n} \exp\left\{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right\}\right) = \left(-\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

これを最大にするパラメーターの条件は

$$\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 \to 最小$$

（<u>最小二乗法</u>）

「同じくらいの誤差ならモデルは単純な方が良い」

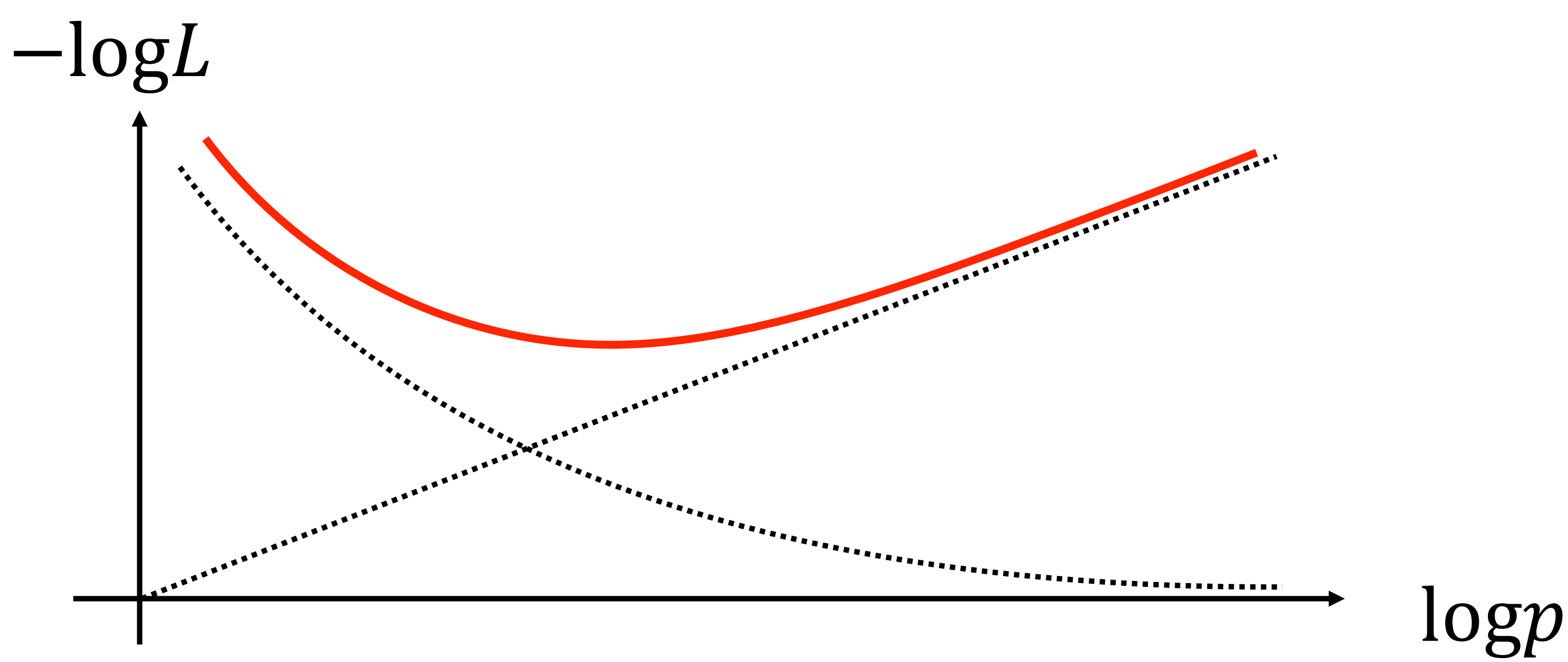
↓

<u>赤池情報量基準</u>

(Akaike Information Criterion, AIC)

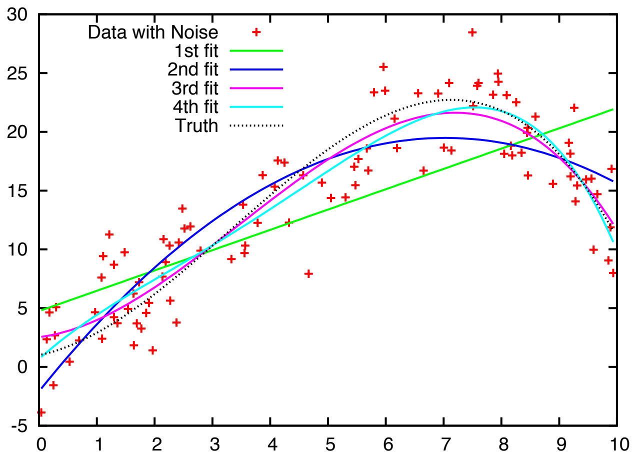
$$\mathrm{AIC} = -2\log L + 2\log p$$

($\log L$:"*最大対数尤度*" ,p:"*パラメター数*")



<u>AICが最も小さくなる p を選択</u>

スパースモデリング



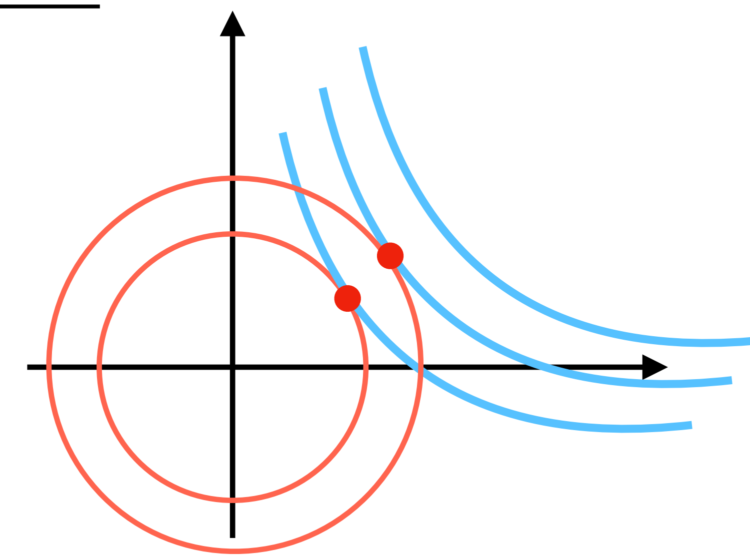
- **線形回帰と最小２乗問題**

$$\vec{y} = X\vec{\beta} + \vec{\xi} \qquad \boxed{\min\left(||y - X\vec{\beta}||_2^2\right)}$$

<u>説明変数(β)は少ない方が良い！</u>

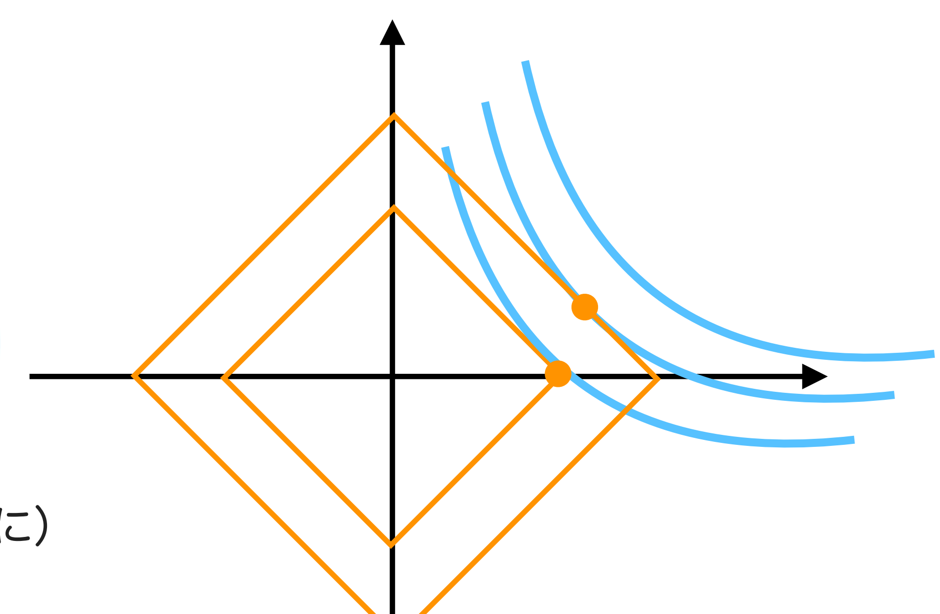
- **Ridge**

$$\min\left(||y - X\vec{\beta}||_2^2 + \boxed{\lambda||\vec{\beta}||_2^2}\right)$$

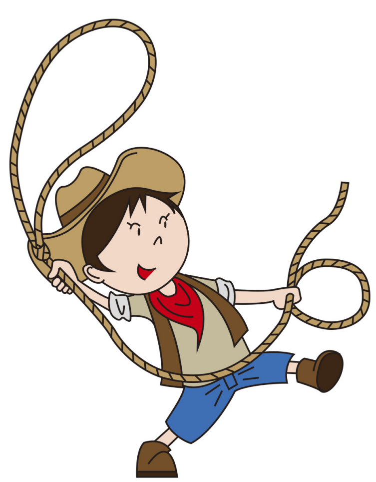


- **LASSO**

$$\min\left(||y - X\vec{\beta}||_2^2 + \boxed{\lambda||\vec{\beta}||_1}\right)$$

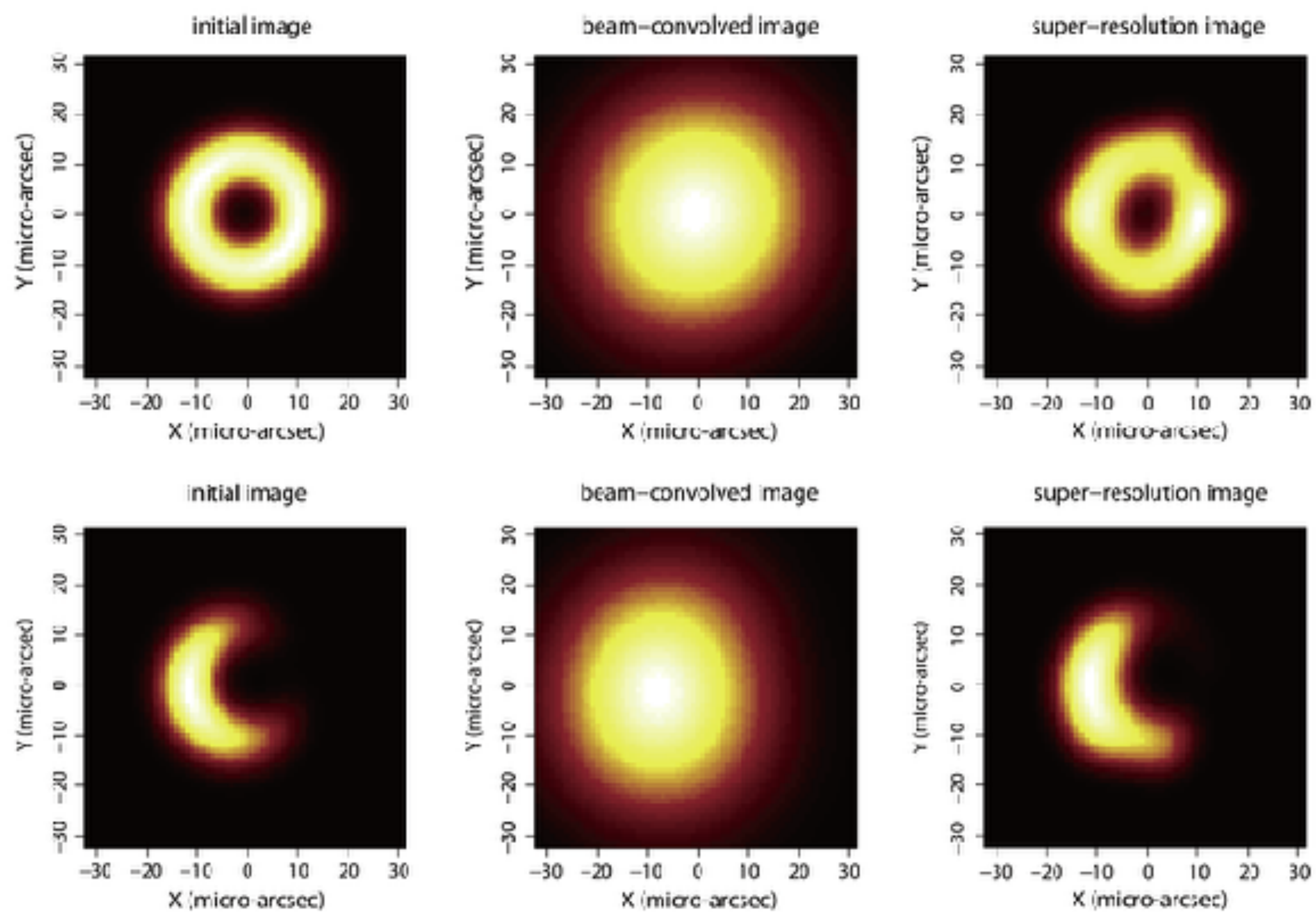


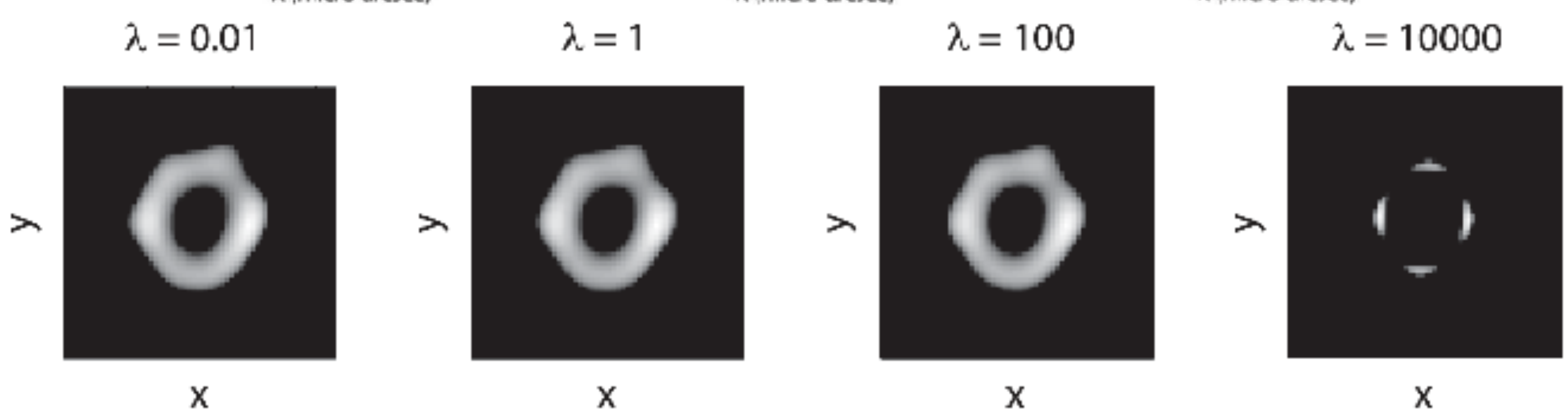
(Tibshirani, Robert (1996) 以降ポピュラーに)



hrukyさんによるイラストACからのイラスト

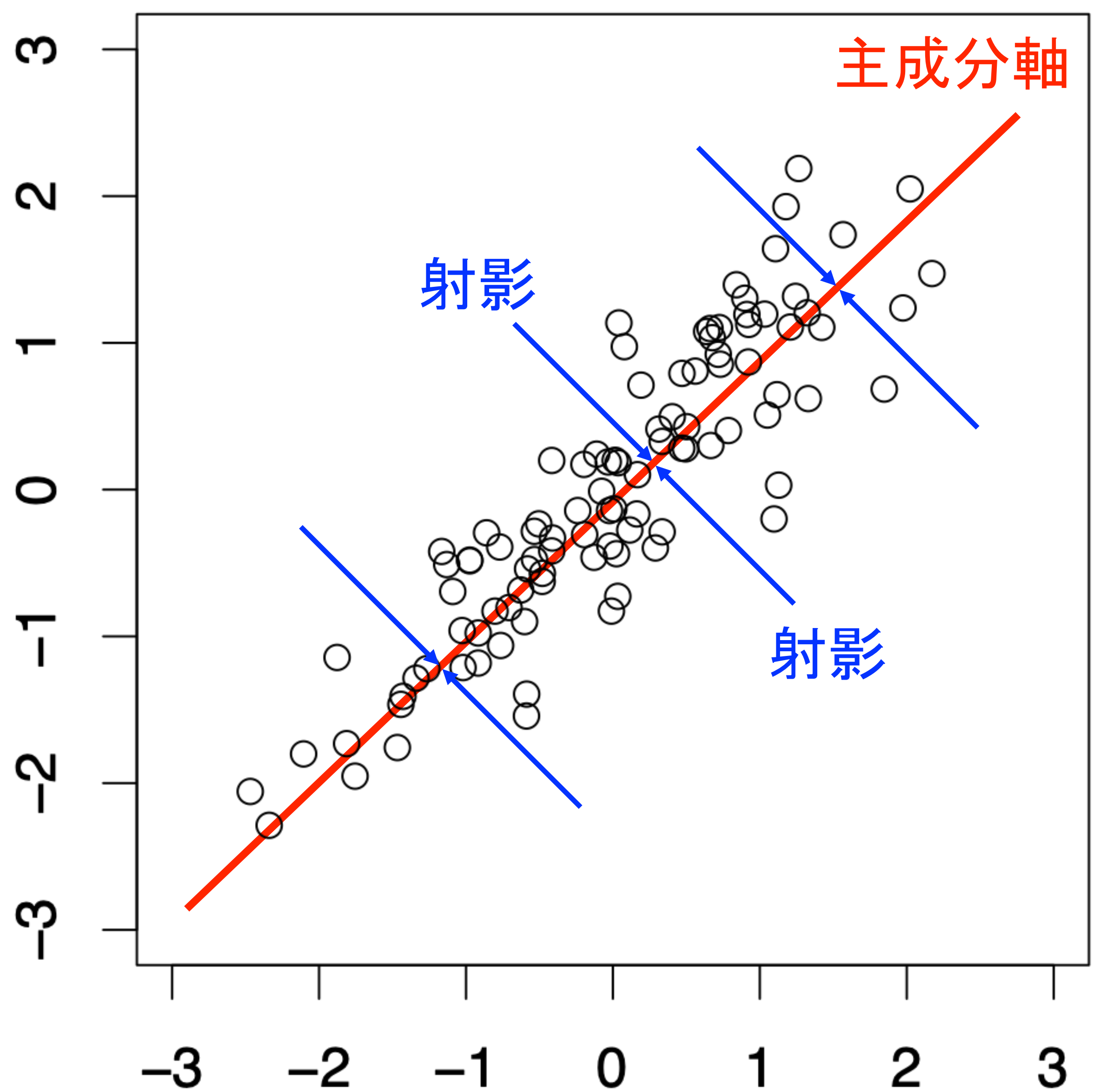
LASSOの天文学への応用例





主成分分析

元のデータの情報をなるべく保存しつつ次元を減らしたい
→「データが一番広がっている部分空間に射影」
→ データの慣性主軸、分散共分散行列の対角化、特異値分解



オートエンコーダ

著作権等の都合により、ここに挿入されていた画像を削除しました。

制限ボルツマン・マシンによる画像再構築の図

G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", Science 28 Jul 2006: Vol. 313, Issue 5786, pp. 504-507 Fig.1.

https://science.sciencemag.org/content/313/5786/504

著作権等の都合により、ここに挿入されていた画像を削除しました。

ロジスティックPCAによる画像再構築の図

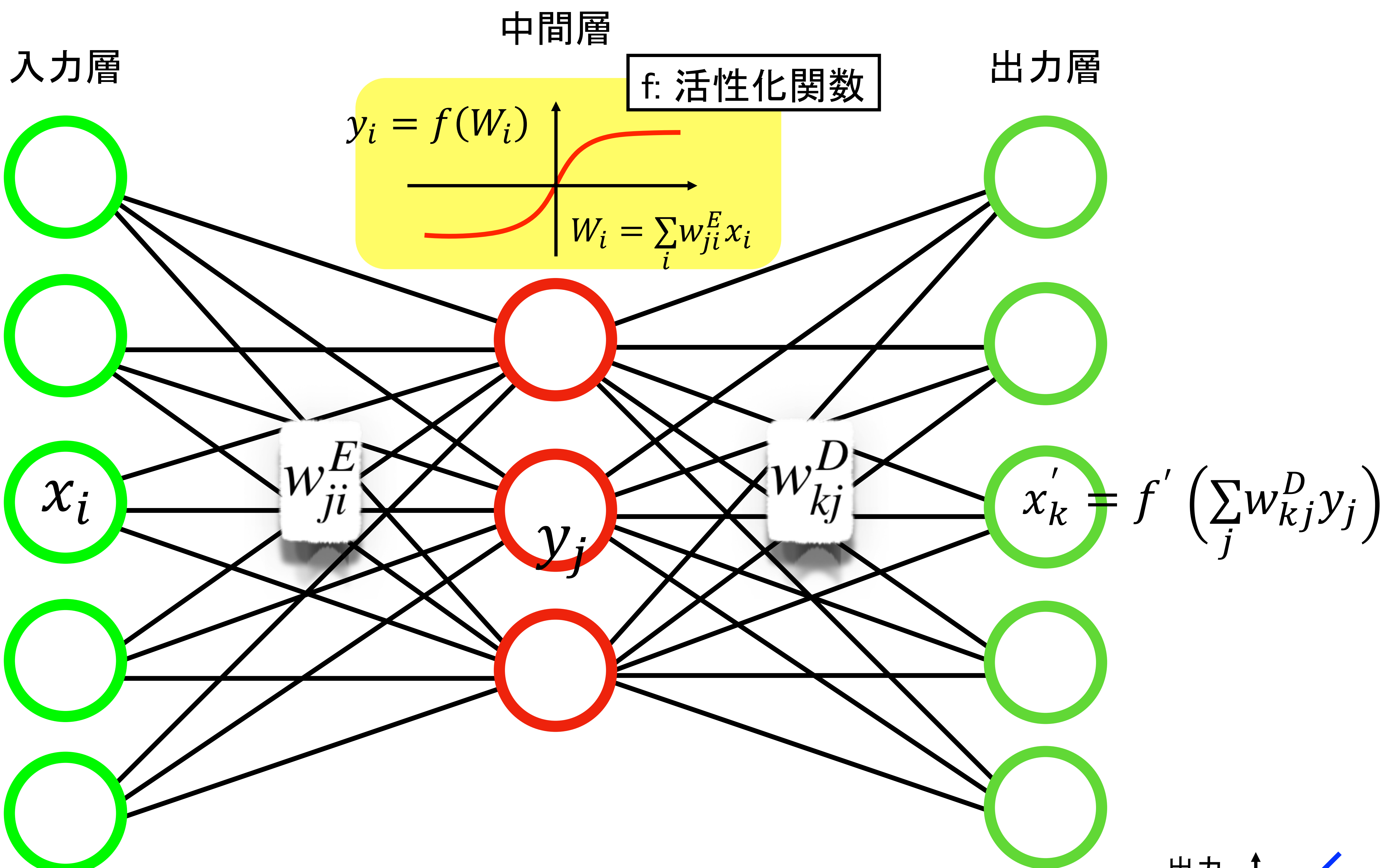
G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", Science 28 Jul 2006: Vol. 313, Issue 5786, pp. 504-507 Fig.3.

https://science.sciencemag.org/content/313/5786/504

（1行目：元画像、2行目：ディープオートエンコーダ、3行目以降：他の次元圧縮法）

"Reducing the Dimensionality of Data with Neural Networks",
G. E. Hinton and R. R. Salakhutdinov, Science 313 (5786), 504–507 (2006).

オートエンコーダと主成分分析



中間層が一層で活性化関数が恒等写像の場合は
主成分分析による次元削減と一緒