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(ピクセル値)の特異値分解

=

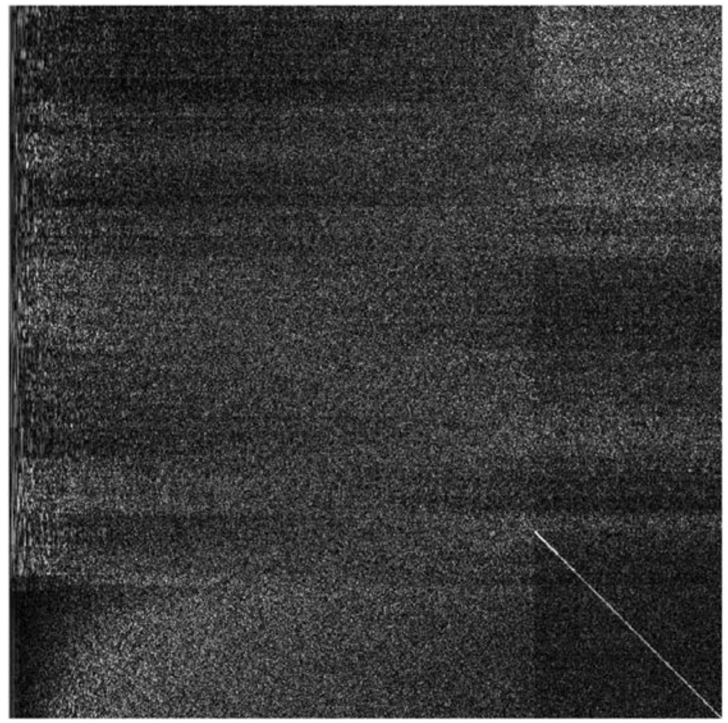
233	232	231	229	227	225	225	225	221	195	181	186	152
233	232	231	229	228	227	227	227	215	172	150	161	138
234	233	231	230	229	229	229	229	216	159	134	108	76
235	234	232	229	229	230	230	228	215	174	149	111	63
235	234	231	229	229	230	228	225	203	184	150	134	84
235	234	231	228	228	229	226	221	190	177	160	136	99
234	233	231	228	228	228	224	218	164	148	147	117	98
233	233	230	227	228	228	224	217	172	144	114	100	93
231	232	228	228	228	225	221	211	172	164	120	76	80
231	231	227	227	227	225	216	196	162	128	120	88	67
229	229	227	228	228	227	213	181	156	120	99	102	82
227	227	228	229	226	227	211	173	144	129	79	99	104
226	228	228	226	221	224	214	181	158	145	107	94	105
224	226	225	223	219	224	223	203	181	160	133	95	95
219	223	223	222	219	221	222	212	180	158	126	92	88
216	222	222	221	217	212	210	203	182	162	131	93	88
215	215	219	220	212	211	201	168	162	151	138	81	80
213	223	220	209	201	200	191	167	155	141	103	54	51
213	218	211	208	205	191	177	167	157	131	79	46	51
204	210	208	207	198	181	179	183	168	141	91	49	47
190	198	202	203	196	191	187	175	165	150	119	76	57
185	187	192	198	191	182	178	167	142	114	95	76	54
175	181	185	187	167	141	149	172	142	97	87	66	49
157	172	173	174	162	132	123	139	128	103	123	78	63
160	165	168	170	158	151	141	109	72	97	128	141	82
156	159	163	166	156	151	148	128	77	72	117	161	113
160	157	155	156	155	157	153	135	91	79	105	147	139
155	158	156	150	149	153	151	138	125	142	120	117	144
156	152	147	144	148	147	138	132	151	151	147	146	135
159	148	140	137	135	122	111	121	147	139	133	135	136
153	152	144	124	104	84	82	109	135	111	104	101	117

画像の特異値分解

A

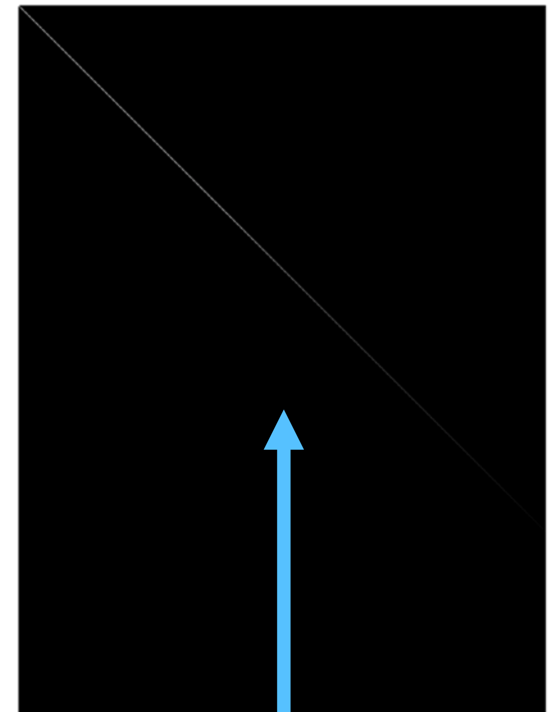
$=$

U

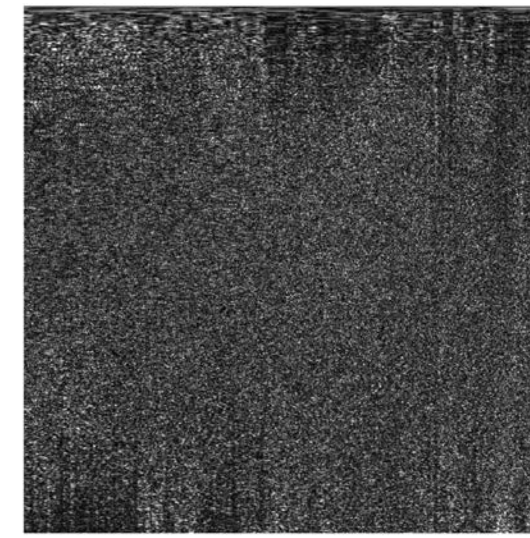


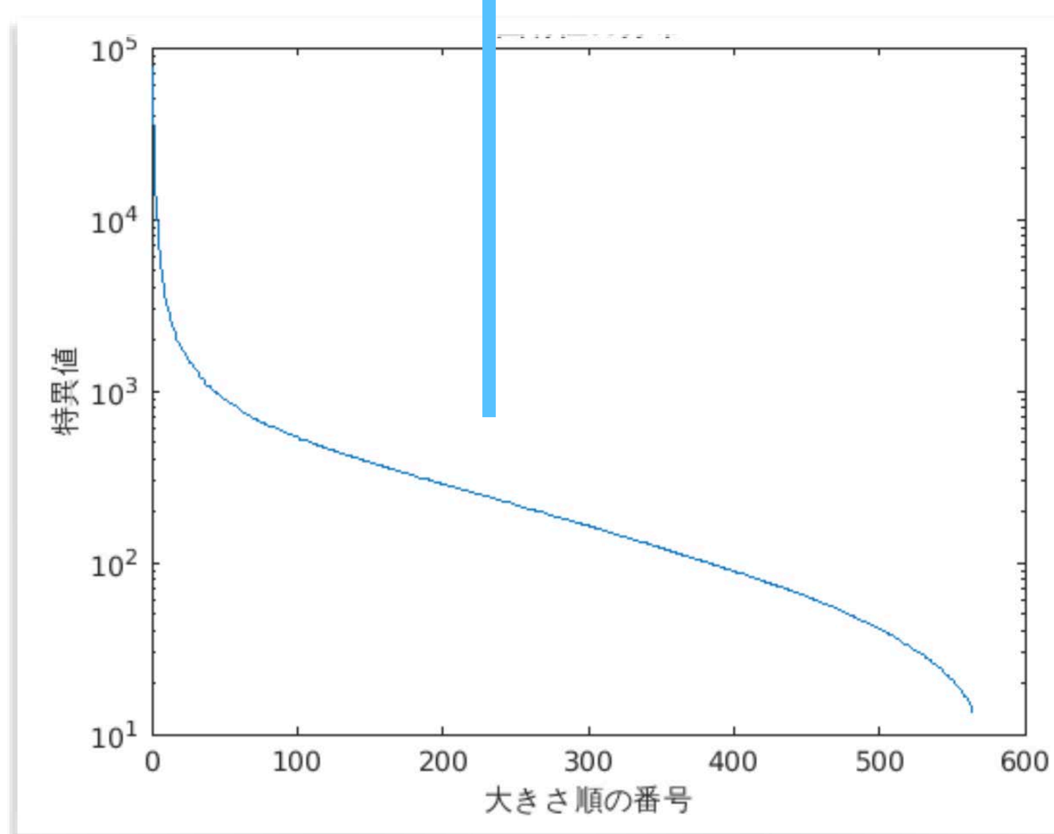
（要素の絶対値で表示）

Σ



$V^\top$



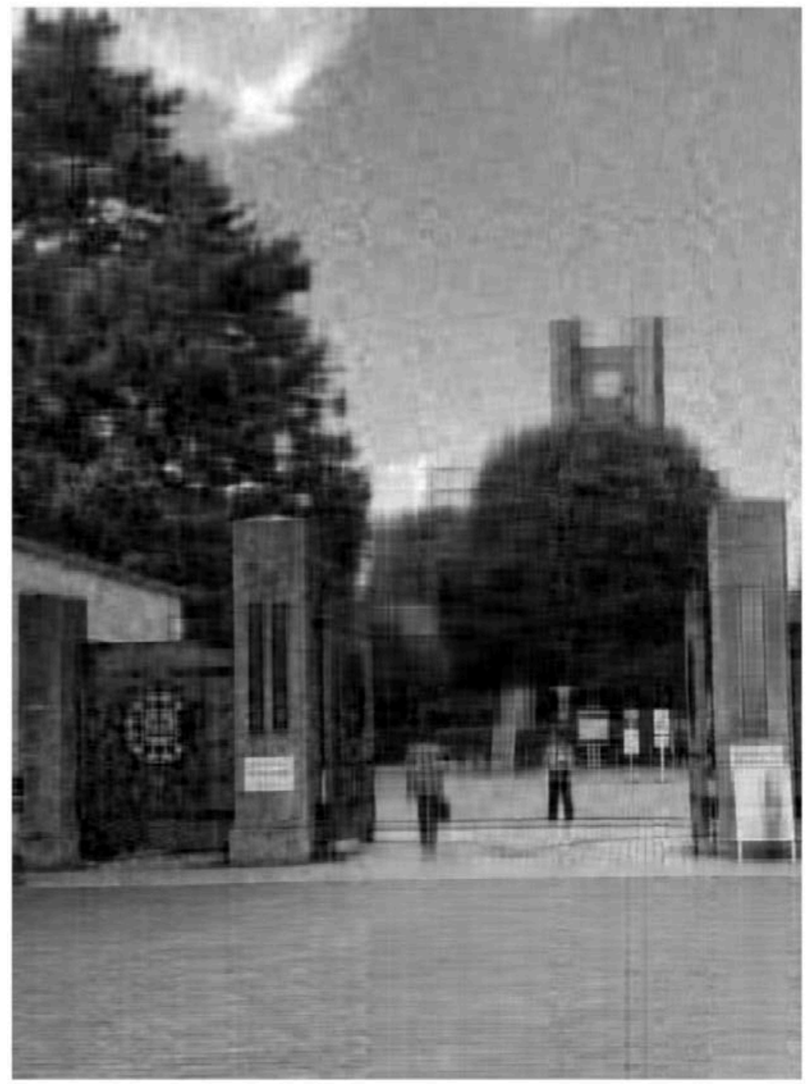


フルランクの特異値をとらなくても、
結構良い再構成ができるのでは？
↓
画像の圧縮
（特異値分解の低ランク近似）

特異値分解の低ランク近似による画像の圧縮

100次元

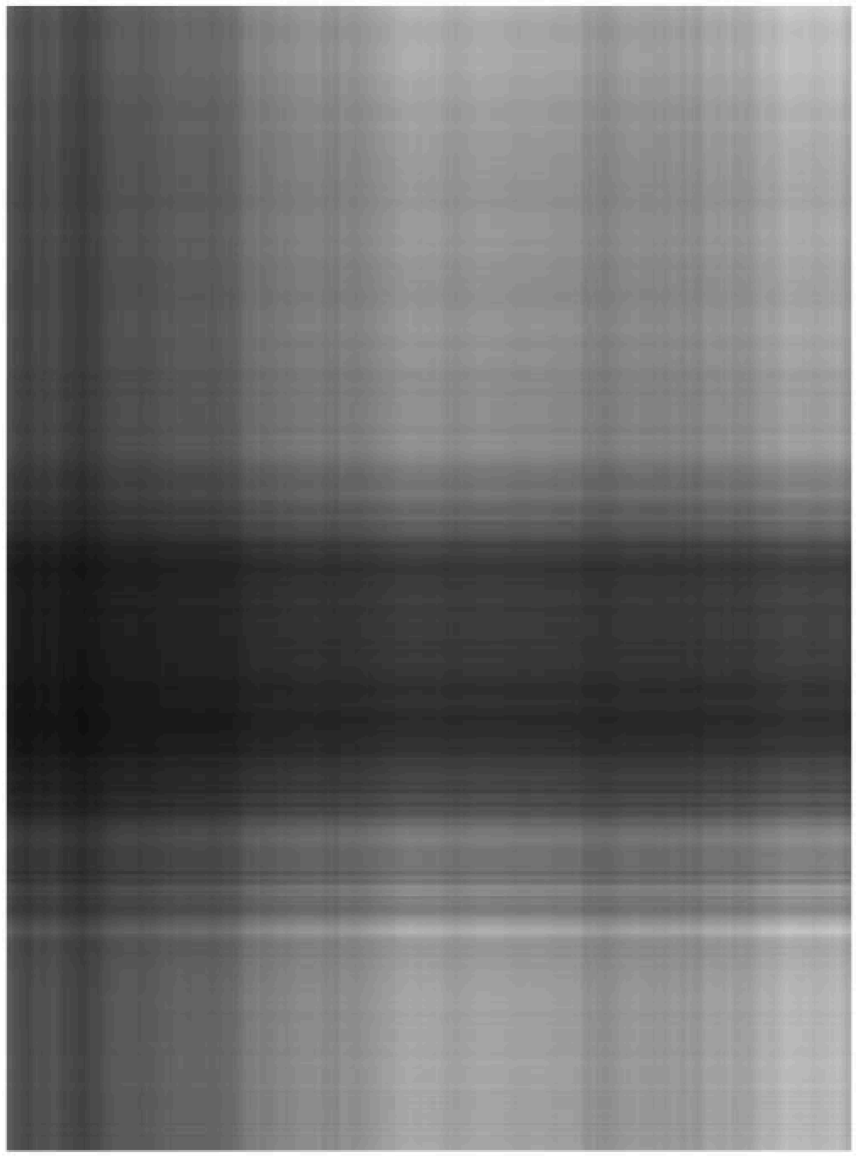
30次元



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

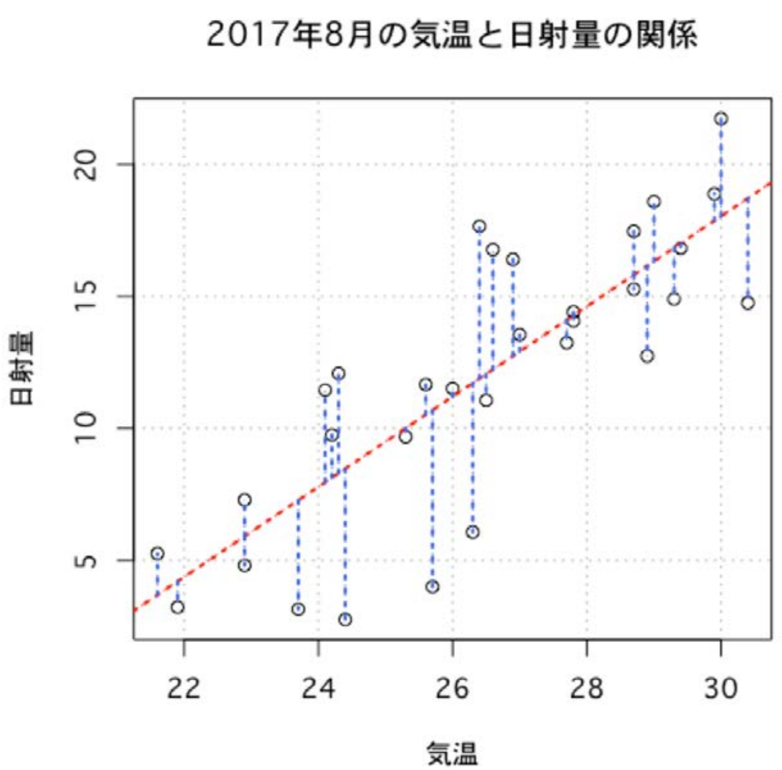
$$\| A\left(\vec{x}_* + \vec{\epsilon}\right) - \vec{b} \|_2^2 - \| A\vec{x}_* - \vec{b} \|_2^2$$

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

↓

<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？

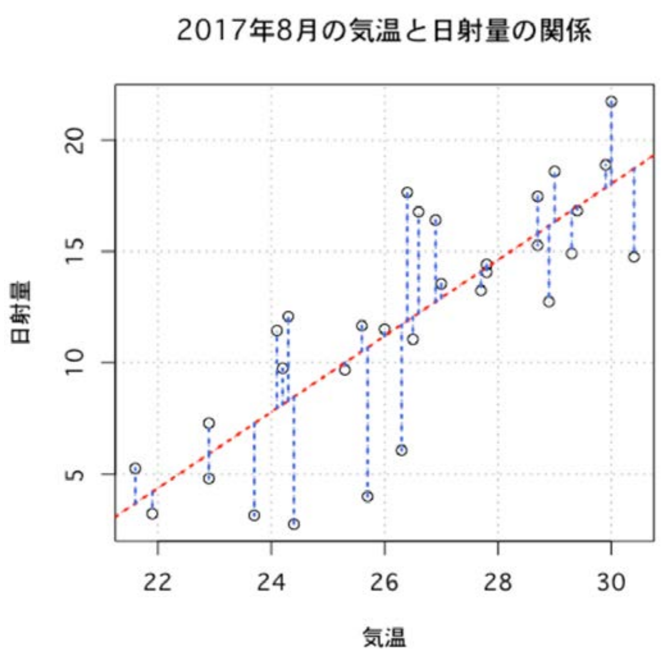
尤度：$L(p) = p^m(1-p)^{n-m}$ *を最大化*

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

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

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

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

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

（<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」

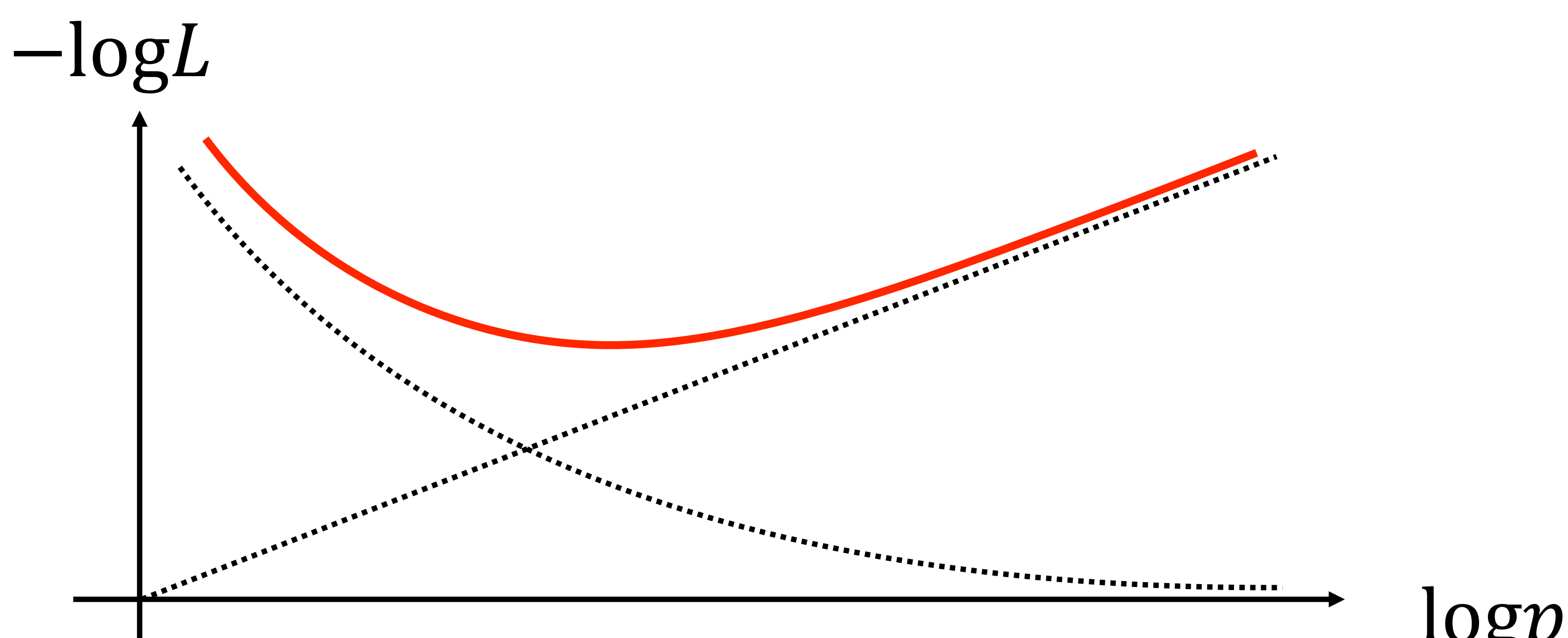
↓

<u>赤池情報量基準</u>

(Akaike Information Criterion, AIC)

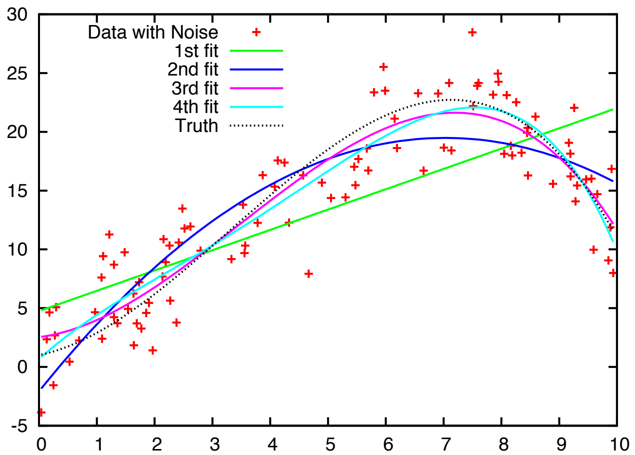
$$\mathrm{AIC} = -2\log L + 2\log p$$

($\log L$:"*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



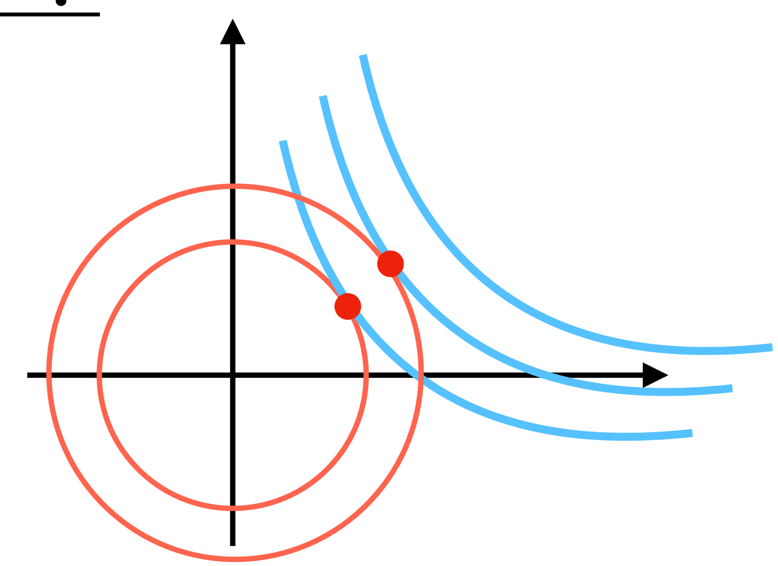
- **線形回帰と最小２乗問題**

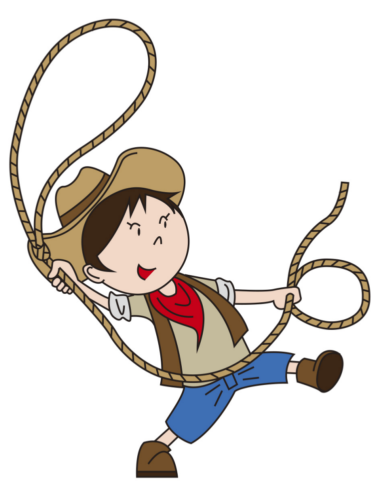
$$\vec{y} = X\vec{\beta} + \vec{\xi} \qquad \boxed{\min\left(||y - X\vec{\beta}||_2^2\right)}$$

<u>説明変数(β)は少ない方が良い！</u>

- **Ridge**

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

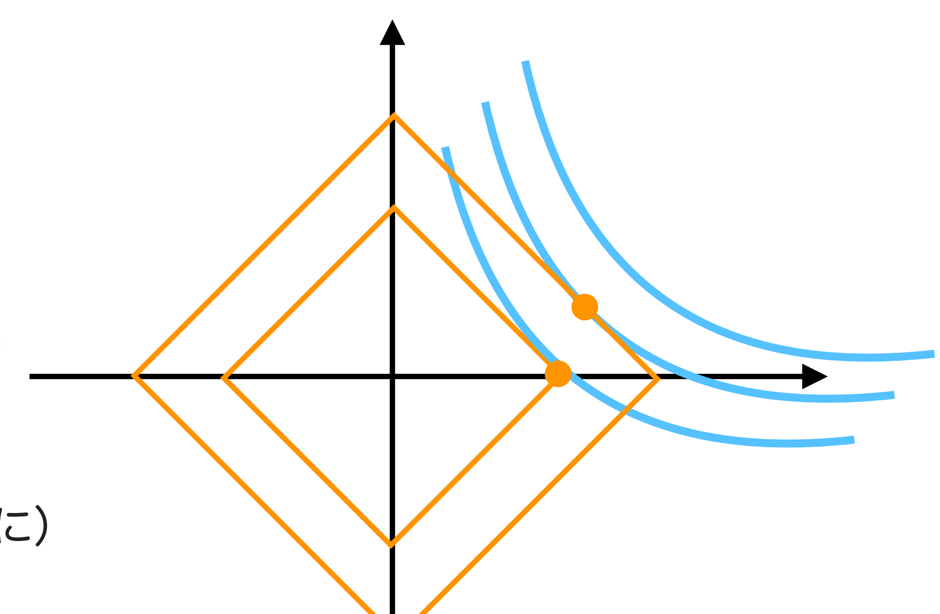




hrukyさんによるイラストACからのイラスト

- **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)

LASSOの天文学への応用例