



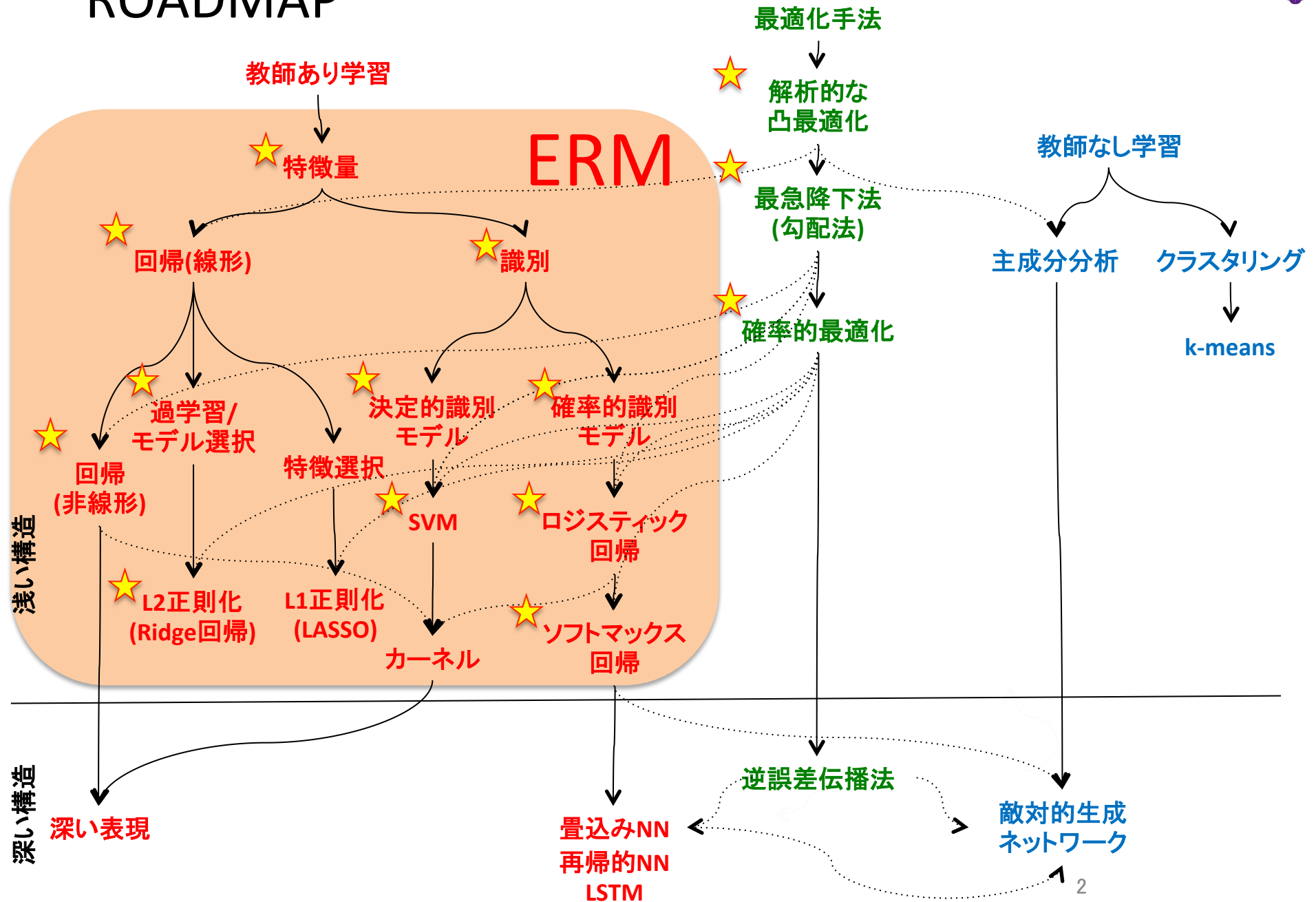
機械学習(8)

k-meansと主成分分析

情報科学類 佐久間 淳



ROADMAP





見通し

	教師あり	教師なし
予測対象が連続	回帰、推薦	PCA
予測対象が離散	分類(識別)	クラスタリング



教師なし学習

- これまでは...教師あり学習をやってきました

- 訓練事例 $\{(x_i, y_i)\}$ から予測機 $f: X \rightarrow Y$ を学習
- 未知事例 $\{x_j\}$ のラベル y_j を予測
- Y が...

$$P(y|x) \text{ に興味アリ}$$

- 連続値 R なら回帰,
- 条件付き確率 $[0,1]$ なら確率的識別モデル(e.g., ロジスティック回帰)
- 離散値なら決定的識別モデルによる分類(k-NNとかSVMとか)

- 教師なし学習では事例 $\{x_i\}$ しか与えられない

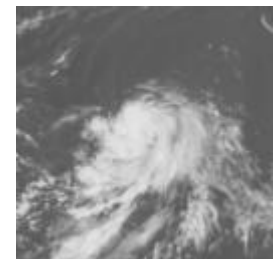
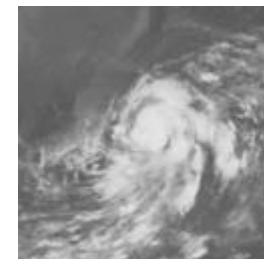
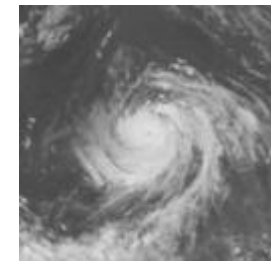
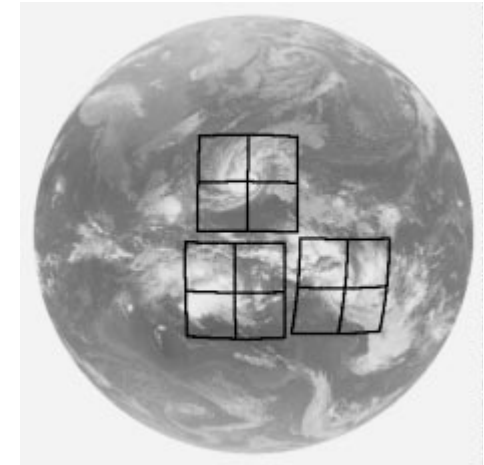
- 事例の集合からわかることを調べる. たとえば...
- 潜在的な変数の推定(クラスタリング, PCA etc)
- 潜在的なパターンの抽出(相関ルール etc)

$$P(x) \text{ に興味アリ}$$



台風雲パターンをクラスタリング

- 台風の雲画像データ
 - データ: 512×512ピクセルの画像
 - データセットのデータ点数: 34000
 - クラスタ数 : 100
- 台風の雲の配置は、どんなパターンがあるか？
 - 過去の台風の分析への利用
 - 未来の台風の動きの予測への利用



引用:

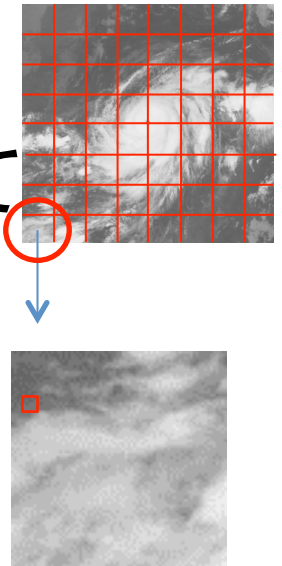
Asanobu Kitamoto , Data Mining for Typhoon Image Collection

<http://agora.ex.nii.ac.jp/digital-typhoon/>



雲画像データ

- 512×512ピクセルの雲画像を8×8ピクセルごとに格子状に区切る
- 各格子内のそれぞれのピクセルについて
 - 曇りなら1
 - 曇りでなければ0
- 各格子内の曇りの割合のベクトルを
n番目のデータの特徴ベクトル x_n とする

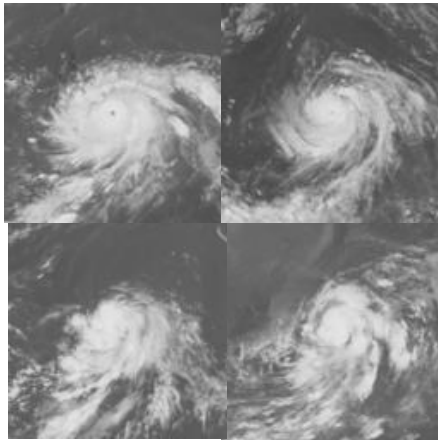


$$\begin{bmatrix} \frac{37}{64} & \cdots & \frac{45}{64} \\ \vdots & \ddots & \vdots \\ \frac{14}{64} & \cdots & \frac{58}{64} \end{bmatrix} \quad \longrightarrow \quad x_n = \begin{pmatrix} \frac{37}{64} \\ \vdots \\ \frac{58}{64} \end{pmatrix}$$



台風のパターンの解析

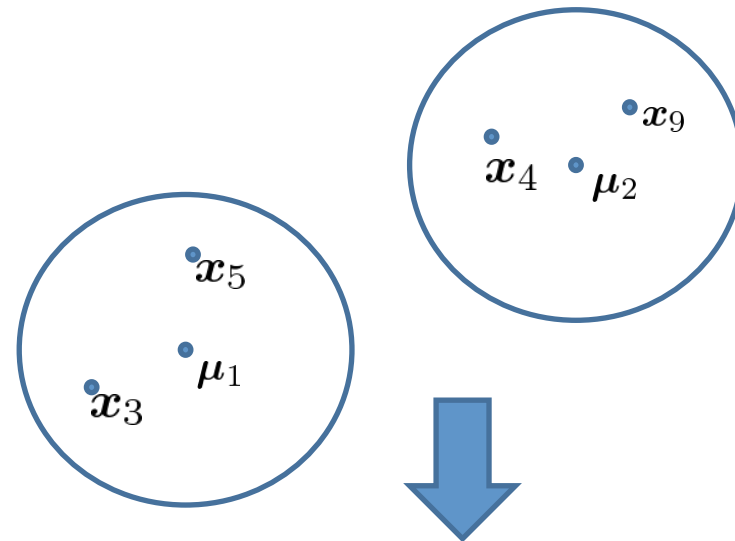
台風の雲の画像



各画像の特徴ベクトルを生成

$$\{x_1, \dots, x_n\}$$

k個のクラスターで分類

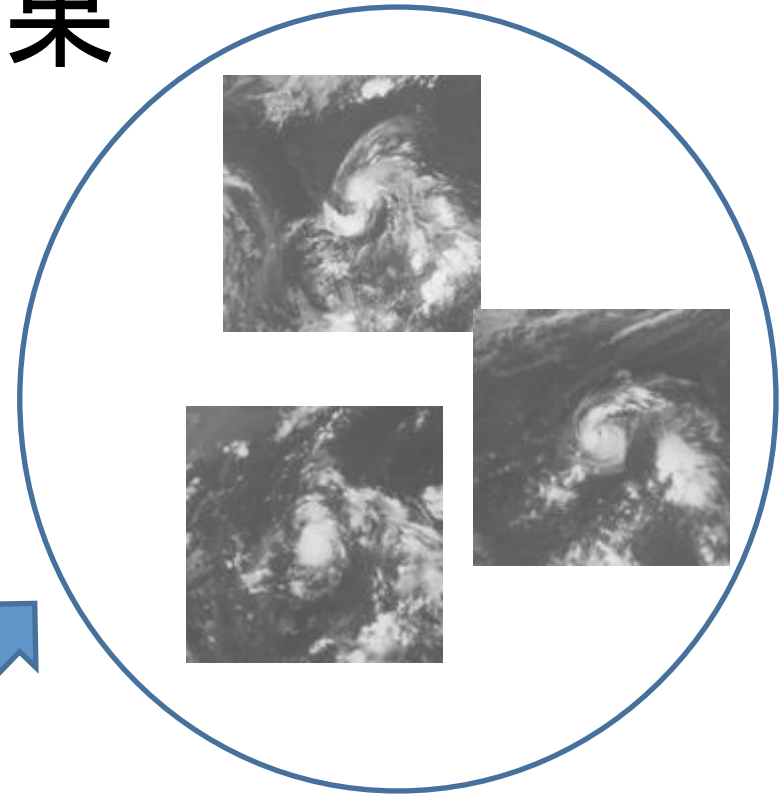
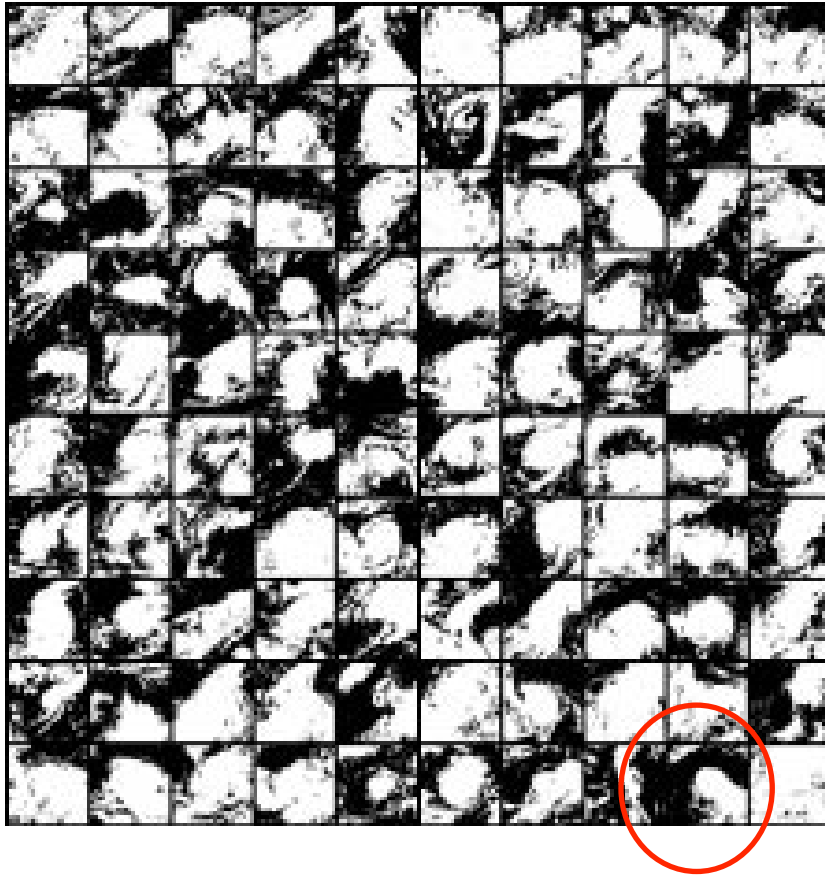


k個のクラスター中心を台風のパターンとして出力

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$$



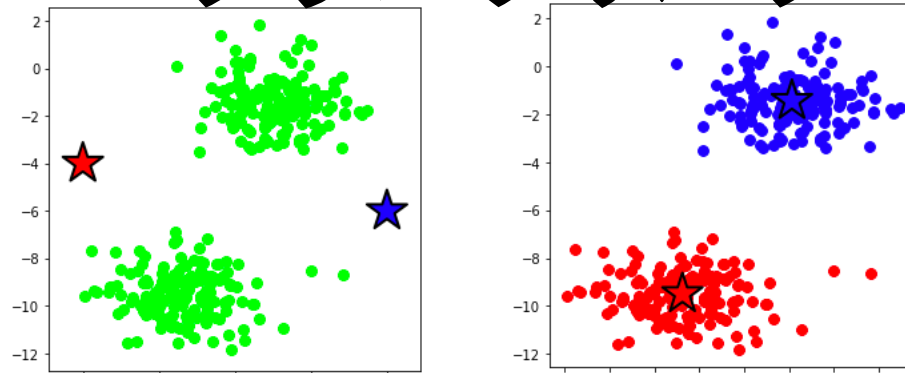
計算結果



- 100種類の台風の雲パターンを得られた
- この結果は似た台風の規模、動きの予測に利用される



クラスタリング



- データを, **教師ラベルなし**で似たもの同士で仲間(クラスタ)分けする方法
 - 同一クラスタ内のデータはより似ている
 - 異なるクラスタ間のデータはより似ていない
 - 似ている...とは? → 距離, あるいは類似度
 - ユークリッド距離 $d(x, y) = (x - y)^T (x - y)$
 - コサイン類似度 $s(x, y) = \frac{x^T y}{\sqrt{(x^T x)(y^T y)}}$
- 両者は何が違う?**



ユークリッド距離のクラスタリング: k-means

- データ: $\{x_1, \dots, x_N\}, x_i \in \mathbb{R}^D$
- プロトタイプ: クラスタの代表点
 - K個のプロトタイプ $\{\mu_1, \dots, \mu_K\}, \mu_j \in \mathbb{R}^D$
- 割当変数 $r_{ik} = \begin{cases} 1 & \text{if } x_i \text{ is assigned to } \mu_k \\ 0 & \text{o.w.} \end{cases}$



k-meansの目的関数

- やはり最適化問題として考える
- 歪み尺度最小化
 - プロトタイプ μ_k に割当てられているデータがどのくらい散らばっているか？

$$J(\underbrace{\mu_1, \dots, \mu_K, r_{11}, \dots, r_{nk}}_{\text{simple}}) = \sum_{k=1}^K \sum_{n=1}^N r_{nk} \|\mathbf{x}_n - \mu_k\|_2^2$$

簡単に $J(\mu_k, r_{nk})$ とかく

K-means clustering $\operatorname{argmin}_{\mu_k, r_{nk}} J(\mu_k, r_{nk})$



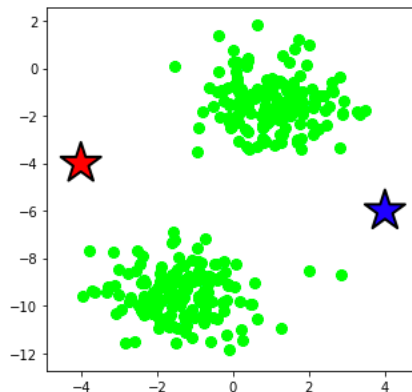
アルゴリズム

1. プロトタイプ μ_k をランダムに初期化 ($k = 1, \dots, K$)

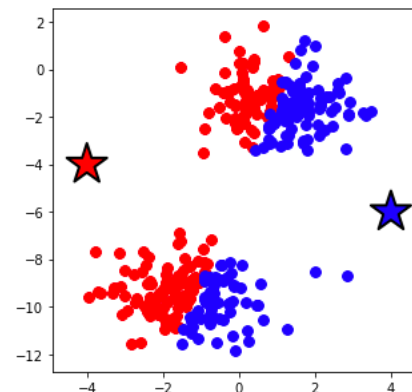
2. r_{nk} を以下の式で決定

$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|_2^2 \\ 0 & \text{otherwise.} \end{cases}$$

Step 1



Step 2



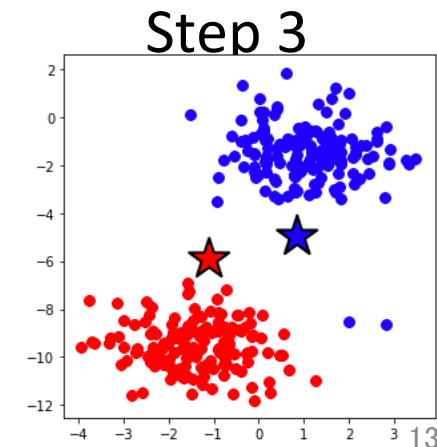
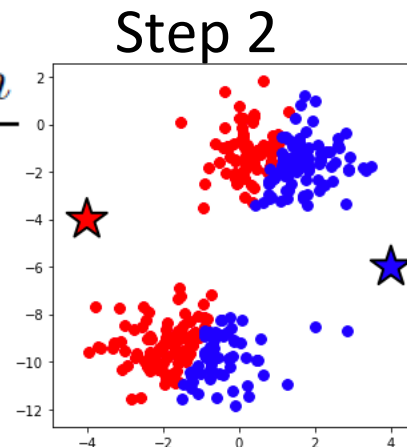


アルゴリズム

1. プロトタイプ μ_k をランダムに初期 ($k = 1, \dots, K$)
2. r_{nk} を以下の式で決定
$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_j \|\mathbf{x}_n - \mu_j\|_2^2 \\ 0 & \text{otherwise.} \end{cases}$$
3. プロトタイプ μ_k を以下の式で決定

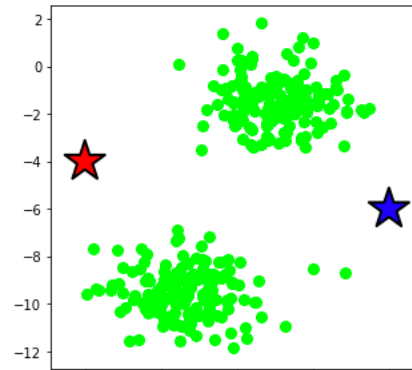
$$\mu_k = \frac{\sum_{n=1}^N r_{nk} \mathbf{x}_n}{\sum_{n=1}^N r_{nk}}$$

4. Jが未収束なら2へ

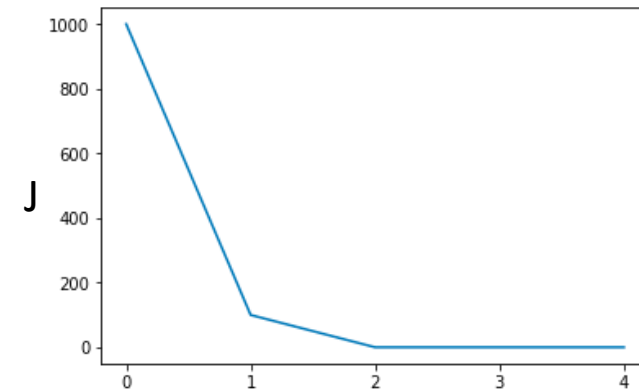




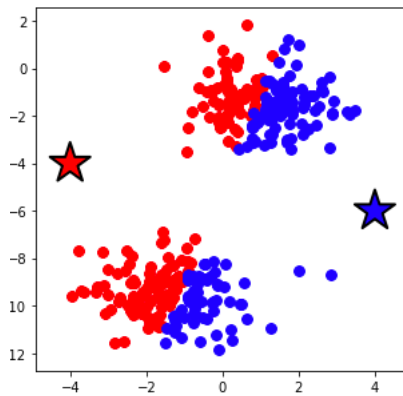
k-meansの挙動



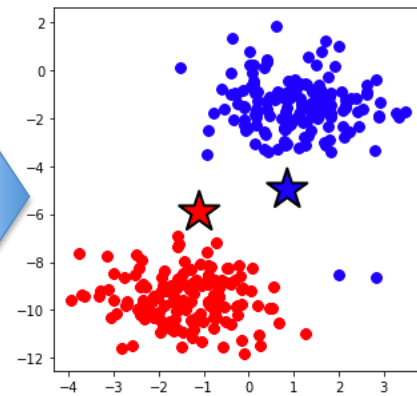
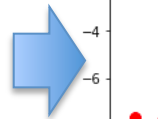
Step 1



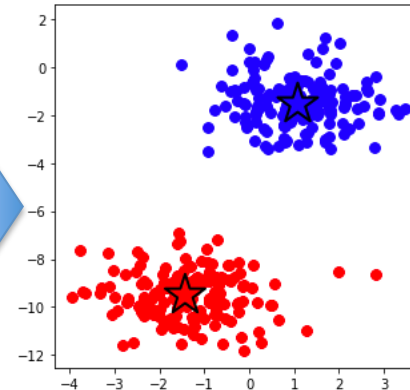
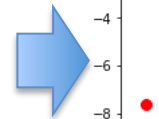
step



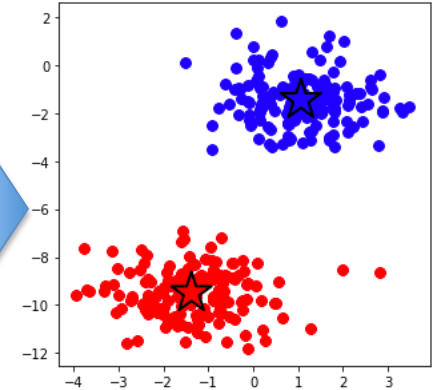
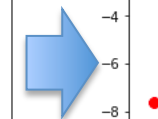
Step 2



Step 3



Step 4



Step 5



Jの性質

- 割当 $r_{nk} \in \{0, 1\}$ は離散変数
 - 勾配を利用した最適化は使えない
 - そもそもは凸関数ではない
- tステップ目の変数を μ_k^t, r_{nk}^t と書くと
 - $J(\mu_k^t, r_{nk}^t)$ は(広義)単調減少する
 - どうやってしめすか？ 演習



7.1 k -means clustering の更新

以下の 4 つのデータを $k = 2$ で k -means クラスタリングすることを考える。

$$\boldsymbol{x}_1 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 5 \\ -4 \end{pmatrix}, \boldsymbol{x}_4 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

ただし初期のクラスタ中心は以下で与えるものとする。

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

以下の問いに答えよ。

1. 初期状態における J を求めよ
2. 1 回目の更新における r_{nk} を求めよ
3. (1) で求めた r_{nk} に基づき、 $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ を更新し、この時の J を求めよ
4. 2 回目の更新では J が変化せず、アルゴリズムが収束することを示せ



7.2 k -means clustering の収束

$\mathbb{R}^D$ のデータ $\mathbf{x}_1, \dots, \mathbf{x}_N$ の k -means クラスタリングを考える。クラスタ数は K , $\mathbf{x}_n$ が k 番目のクラスタに割り当てられていることを示す変数を r_{nk} と表記する。 k 番目のクラスタの中心点を $\boldsymbol{\mu}_k$ とする。 k -means クラスタリングの目的関数を以下のように定義する。

$$J(\boldsymbol{\mu}_k, r_{nk}) = \sum_{k=1}^K \sum_{n=1}^N r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\| \quad (8)$$

$\boldsymbol{\mu}_k$ および r_{nk} を k -means クラスタリングアルゴリズムで更新した時について、以下の問いに答えよ。

1. 時刻 t のステップ 2 において, $J(\boldsymbol{\mu}_k^t, r_{nk}^t) \geq J(\boldsymbol{\mu}_k^t, r_{nk}^{t+1})$ を示せ
2. 時刻 t のステップ 2 において割り当て変数が変更された時, ステップ 3 において $J(\boldsymbol{\mu}_k^t, r_{nk}^{t+1}) \geq J(\boldsymbol{\mu}_k^{t+1}, r_{nk}^{t+1})$ を示せ



7.3 k -means clustering の挙動

プログラム `k_means.ipynb` を実行して以下の問題に答えよ。プログラムは三つの異なる種類のデータについて $k = 3$ の k -means クラスタリングを実行した結果を表している。同じ色のデータ点は同じクラスに割り当てられたことを表している。以下の問いに答えよ。

1. 2 番目のデータは 1 番目のデータに比べてクラスタリングが正しく実行できているとは言えない。1 番目のデータと 2 番目のデータの違いは何か。またその違いに着目して、2 番目のデータのクラスタリングが正しく実行されない理由について考察しなさい。
2. 3 番目のデータは 1 番目のデータに比べてクラスタリングが正しく実行できているとは言えない。1 番目のデータと 3 番目のデータの違いは何か。またその違いに着目して、3 番目のデータのクラスタリングが正しく実行されない理由について考察しなさい。

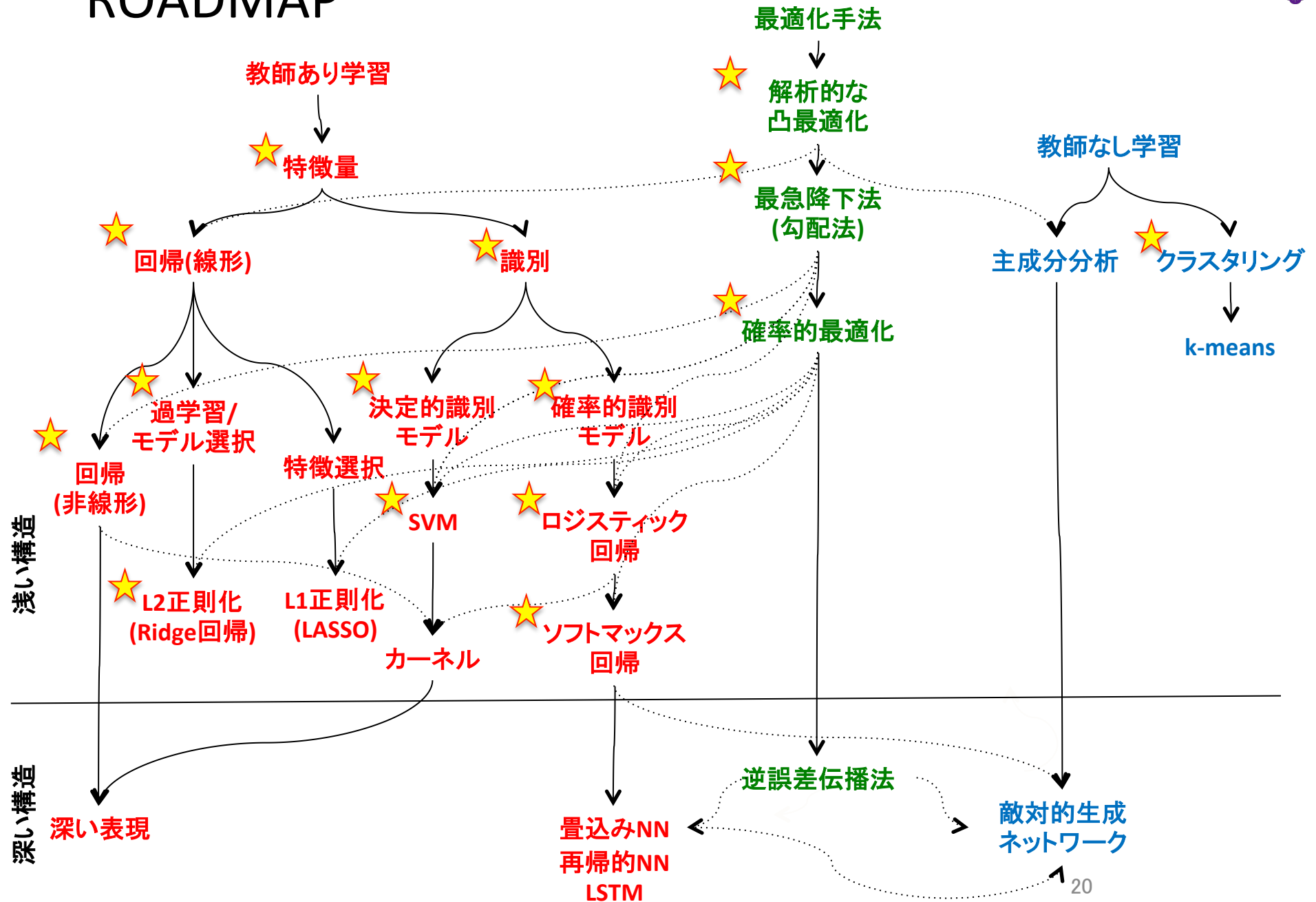


K-meansまとめ

- 複数のベクトルデータを k 個に仲間分け
- クラスタ中心とクラスタ割当を交互に更新
 - J の単調増加性を保証



ROADMAP





見通し

	教師あり	教師なし
予測対象が連続	回帰、推薦	PCA
予測対象が離散	分類(識別)	クラスタリング



偽札の真贋鑑定

- スイス銀行紙幣データ
 - 本物と偽物の1000スイスフラン紙幣のデータ
 - それぞれ100枚
 - 横幅長, 左縦幅長, 右縦幅長, 下枠内長, 上枠内長, 対角長 (6変数)
 - 真札, 偽札の鑑別結果



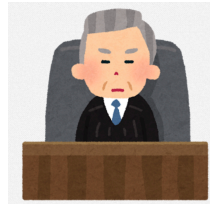
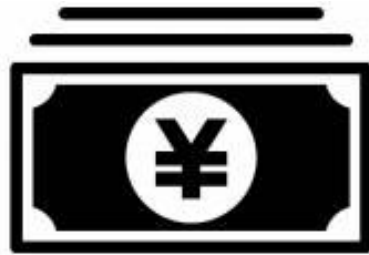


主成分分析

- モデリングの観点からは特徴数が多いほうが分析には有利
- 一方, 特徴数が多くなると人間には一覧できず、データの性質の把握が難しくなる
 - D 特徴ある場合には相関は $D(D-1)/2$ 個
- 主成分分析
 - 多数の特徴量を持つデータについて, 特徴間の相関を排除し, できるだけ少ない情報の損失で少数個の無相関な**合成変数**に縮約し分析する
 - つまり、「次元削減」



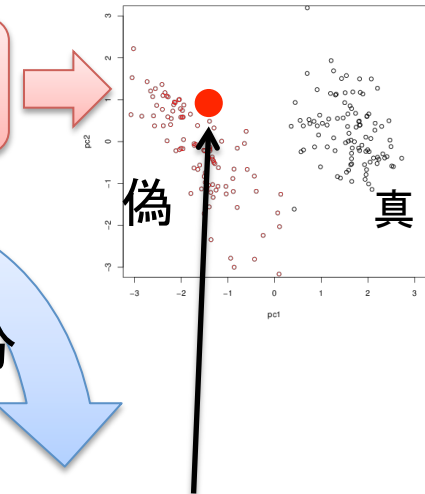
偽札の真贋鑑定(シナリオ)



測定

潜在的な低次元構造の発見

主成分分析



真札
偽札

(7.4, 0.7, 0, 1.9, ...)
(4.0, 5, 7.8, 0.8, ...)

鑑別



測定 (0.76, 0.04, 2.3, ... 4.8)



真贋不明

(-1, -2.1)

贋札と判別



https://www.irasutoya.com/2017/03/blog-post_894.html

https://www.irasutoya.com/2013/07/blog-post_42.html

https://www.irasutoya.com/2018/09/blog-post_781.html

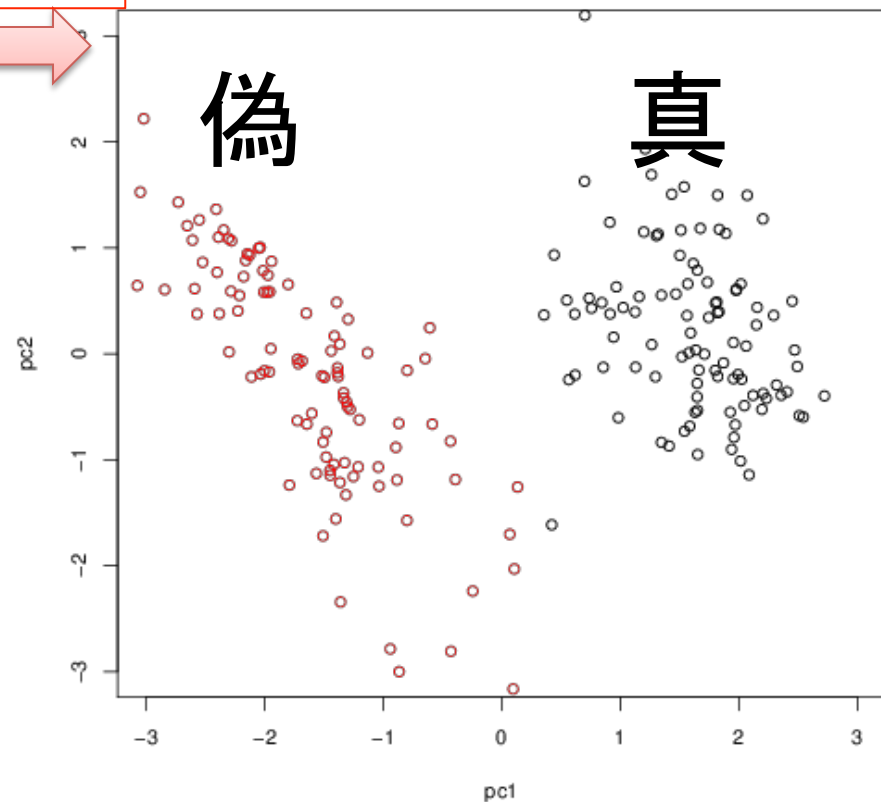
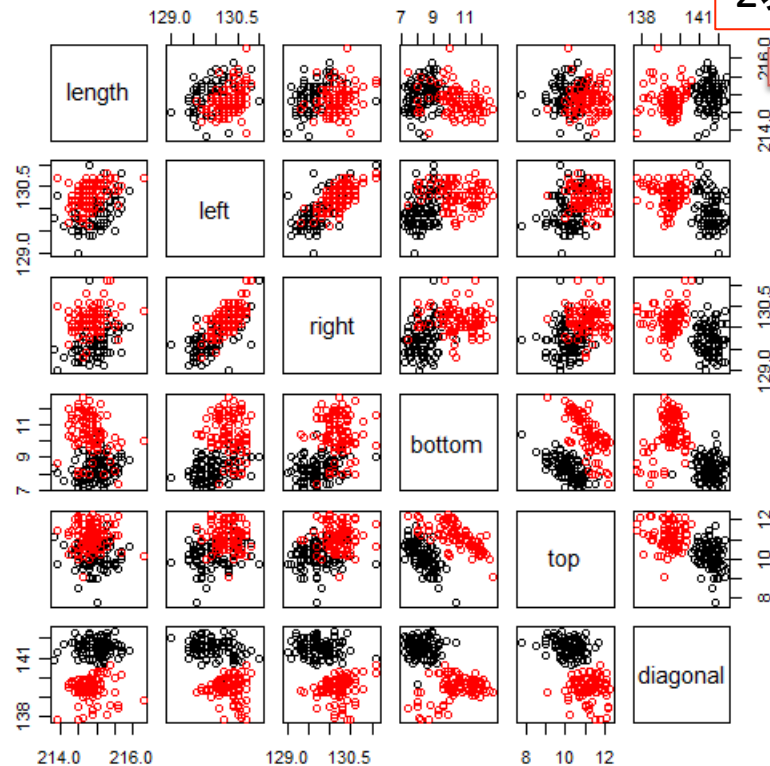
<https://www.silhouette-ac.com/detail.html?id=122085&sw=%E6%9C%AD>



主成分分析→重要な低次元構造への射影

- 属性数が多いほうが有利
- しかし, 属性数が多い(この例では6つ)と人間には解釈が難しい
- 主成分分析を用いる

6次元から
2次元へ射影



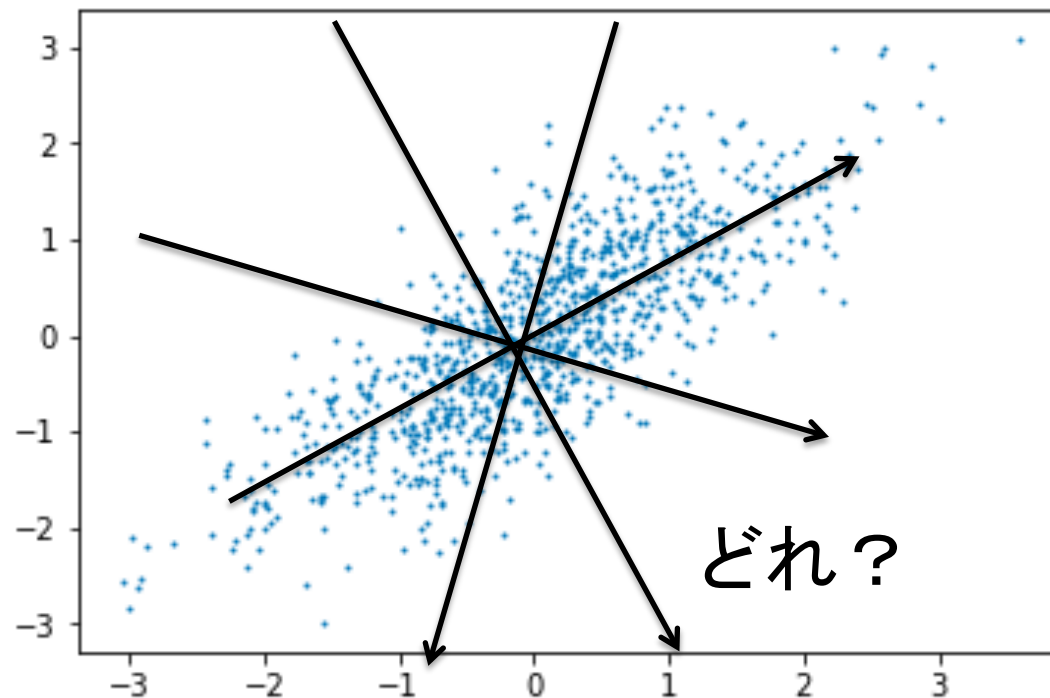


教師あり学習と何が違うのか

- 教師あり学習(分類)
 - 事例 (x_i, y_i) where $x_i \in \mathbb{R}^D, y_i \in \mathcal{Y}$
 - 分類器 $f : \mathbb{R}^D \rightarrow \mathcal{Y}$
 - ラベルをうまく予測する分類器を求めるのが目標
 - $P(y|x)$ に興味がある
- 教師なし学習(PCA)
 - 事例 (x_i) where $x_i \in \mathbb{R}^D$
 - 事例をうまく説明する低次元空間 $\mathbb{R}^M, M < D$ を求めるのが目標
 - 低次元空間での $P(x)$ に興味がある
 - 良い $P(x)$ が求まれば, ラベルは目で見てわかるはず



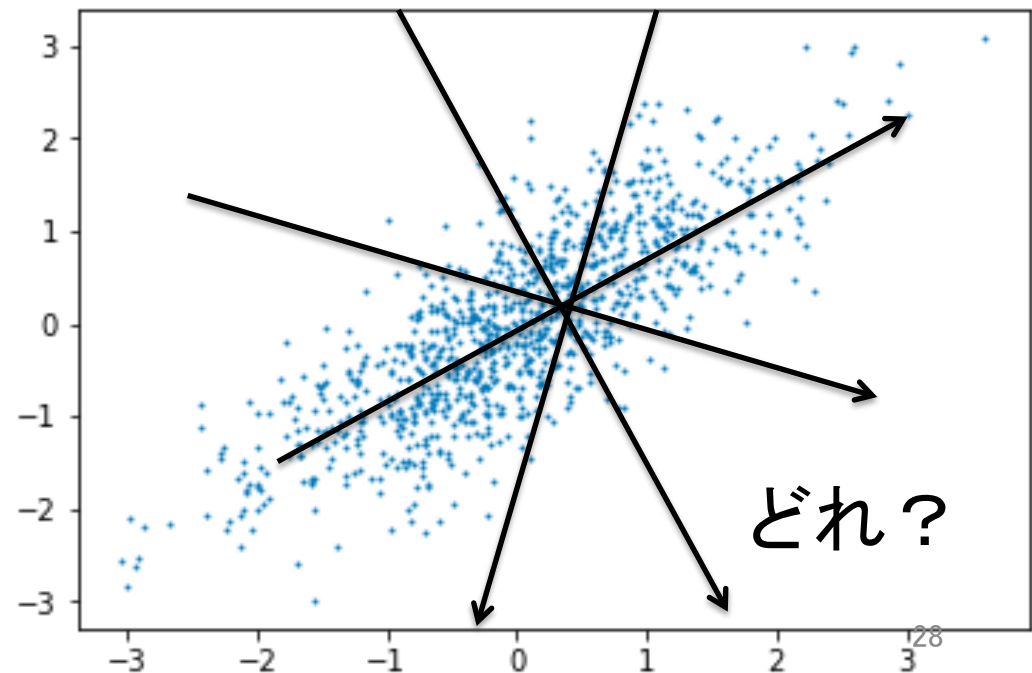
サンプル群を1次元に削減して表現しなければならなかったら、どの方向が最も合理的か？





「情報損失の少ない少数個の 無相関な合成変数」とは？

- サンプルを実際の次元数より小さい次元数を持つ部分空間に射影
- 分散が最大になる方向に射影すればよい





データ

- 測定値 $(\mathbf{x}_n^*)^T = (x_{n1}^*, \dots, x_{nD}^*) \in \mathbb{R}^D$
- $\bar{x}_d = (x_{1d}^* + x_{2d}^* + \dots + x_{Nd}^*)/N$ として
 $x_{nd} = x_{nd}^* - \bar{x}_d$ (中央化, 白色化ともいう)
- 中央化されたデータ

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{ND} \end{pmatrix}$$

$\mathbf{x}_1^T$ →



正規直交基底系

- 正規直交基底系 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_D\}$

$$\begin{cases} \mathbf{u}_i^T \mathbf{u}_i = 1 & \forall i \\ \mathbf{u}_i^T \mathbf{u}_j = 0 & (i \neq j) \end{cases}$$

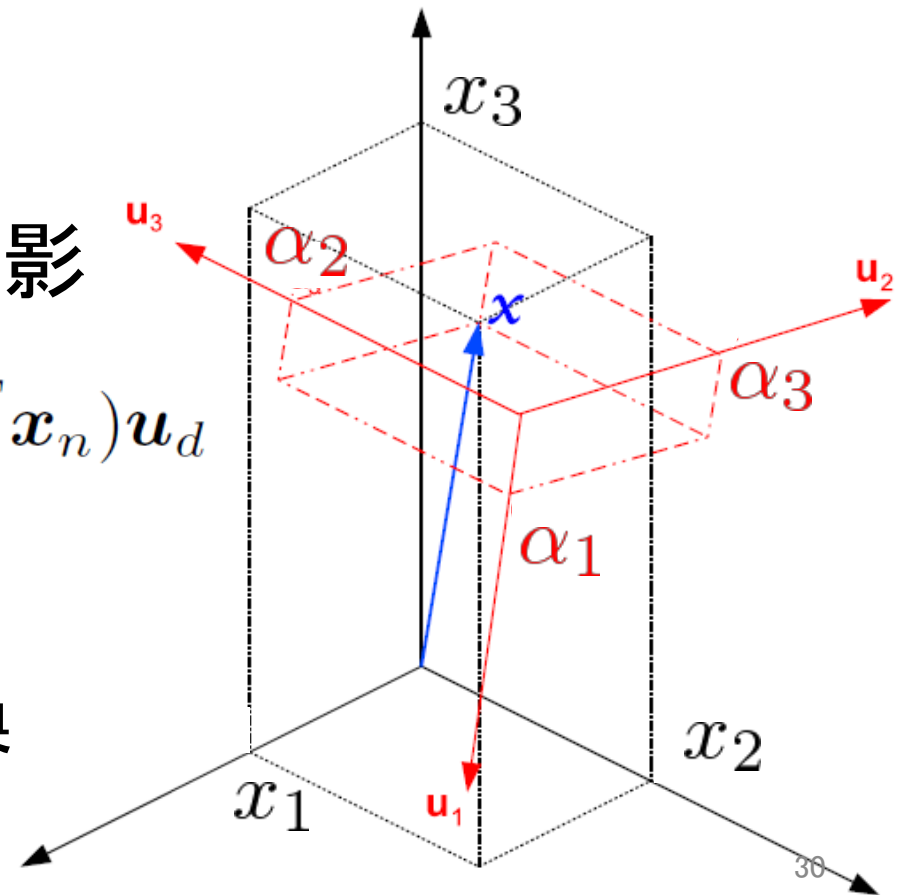
- $\mathbf{x}_n$ の $\{\mathbf{u}_i\}_{d=1}^D$ への射影

$$\mathbf{x}_n = \sum_{d=1}^D \alpha_{nd} \mathbf{u}_d = \sum_{d=1}^D (\mathbf{u}_d^T \mathbf{x}_n) \mathbf{u}_d$$

$$\mathbf{x}^T = (x_1, x_2, x_3)$$


 $(\alpha_1, \alpha_2, \alpha_3)$

基底変換
(演習)





7.4 正規直交基底

$\mathbb{R}^3$ 上の基底系

$$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \right\} \quad (9)$$

を考える。

1. $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ が正規直交基底系であることを示しなさい
2. $\mathbf{u}_1, \mathbf{u}_2$ からなる二次元の正規直交基底系を考える。 $\mathbf{x}^T = (0, \sqrt{2}, \sqrt{3})$ をこの正規直交基底系に射影し、その座標を求めなさい



主成分分析の動機

- D特徴からなるサンプルをM(<D)次元正規直交基底系 $\{u_i\}_{i=1}^M$ に射影することを考える

m番目の基底への射影 $z_m = u_m^T x$

- 問題(M=1の場合) $\{x_1, \dots, x_N\}$ の u_1 への射影を $\{z_{11}, \dots, z_{N1}\}$ としたとき, $\{z_{11}, \dots, z_{N1}\}$ の分散を最大にする u_1 方向をどうやって求めるか？

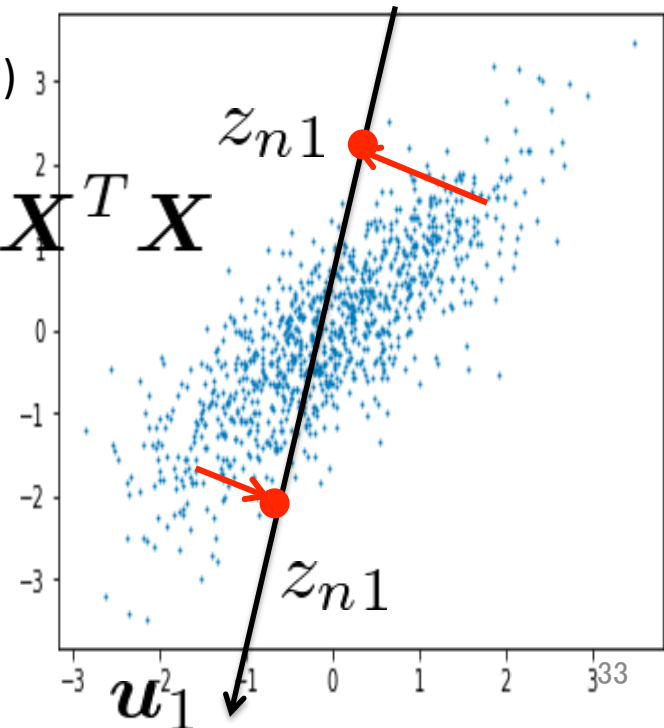


第一主成分について

- 第一主成分得点 $z_1^T = (z_{11}, \dots, z_{N1})$
- 第一主成分に射影した後の成分 $z_1 = X u_1$ (演習)
- 第一主成分得点の平均 $\bar{z}_1 = 0$ (演習)  X は中央化されていることに注意
- 第一主成分得点の分散 (演習)

$$\sigma_{z_1}^2 = u_1^T \Sigma u_1 \text{ where } \Sigma = \frac{1}{N} X^T X$$

$\sigma_{z_1}^2$ を最大にする u_1 を求める
→最適化に帰着





第一主成分を求める最適化

$$\text{maximize } \underbrace{u_1^T \Sigma u_1}_{\text{u}_1 \text{ 方向にサンプルを射影した時の分散}} \text{ subject to } \underbrace{u_1^T u_1 = 1}_{\text{u}_1 \text{ を単位ベクトルとして扱うための制約}}$$

u_1 方向にサンプルを射影した時の分散

u_1 を単位ベクトルとして扱うための制約

- Σ は共分散行列なので半正定値行列
- よって $u_1^T \Sigma u_1$ は凸関数
- 等式制約をもつ凸関数最適化



7.5 主成分の導出 (1)

$\mathbf{x}_i \in \mathbb{R}^D$ ($i = 1, \dots, N$), $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}$ とする。 $\mathbf{x}_1, \dots, \mathbf{x}_N$ の第一主成分を求めるための準備として以下を導きなさい。

1. $\mathbf{u}_1 \in \mathbb{R}^D$ とする。 $\mathbf{x}_1, \dots, \mathbf{x}_N$ をそれぞれ $\mathbf{u}_1$ に射影した座標を $\mathbf{z}_1^T = (z_{11}, \dots, z_{1N})$ とする。 $\mathbf{z}_1 = \mathbf{X}\mathbf{u}_1$ であることを示せ。
2. $\mathbf{x}_1, \dots, \mathbf{x}_N$ が中央化されているならば、 $\bar{z}_1 = \frac{1}{N} \sum_{i=1}^N z_{1i} = 0$ であることを示せ
3. $z_{11}, \dots, z_{1N}$ の分散は $\sigma_{z_1}^2 = \mathbf{u}_1^T \mathbf{\Sigma} \mathbf{u}_1$ であることを示せ。ここで $\mathbf{\Sigma} = \frac{1}{N} \mathbf{X}^T \mathbf{X}$ である
4. $\mathbf{X}$ が中央化されているとき、 $\mathbf{\Sigma} = \frac{1}{N} \mathbf{X}^T \mathbf{X}$ は $\mathbf{x}_1, \dots, \mathbf{x}_N$ の分散共分散行列であることを示せ



凸計画問題

- 凸計画問題
 - 目的関数が凸関数
 - 実行可能領域が凸集合

凸計画問題	制約なし	等式制約	不等式制約
fが微分可能	fの微分を0にする点 が解析的に求まる ⇒その点が最適解 求まらない ⇒最急降下法など	ラグランジュの未 定乗数法	ラグランジュの未 定乗数法 今回は扱わない (SVM導出などで出 てくる)
fが微分不可能	2,3週後に触れます	今回は扱わない	今回は扱わない



ラグランジュの未定乗数法

等式制約がある場合

- 問題 maximize $f(\boldsymbol{x})$ subject to $g(\boldsymbol{x}) = 0$
 - ラグランジュ乗数 λ
 - ラグランジュ関数 $L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})$

- $\boldsymbol{x}$ が(局所)最適解であるならば:

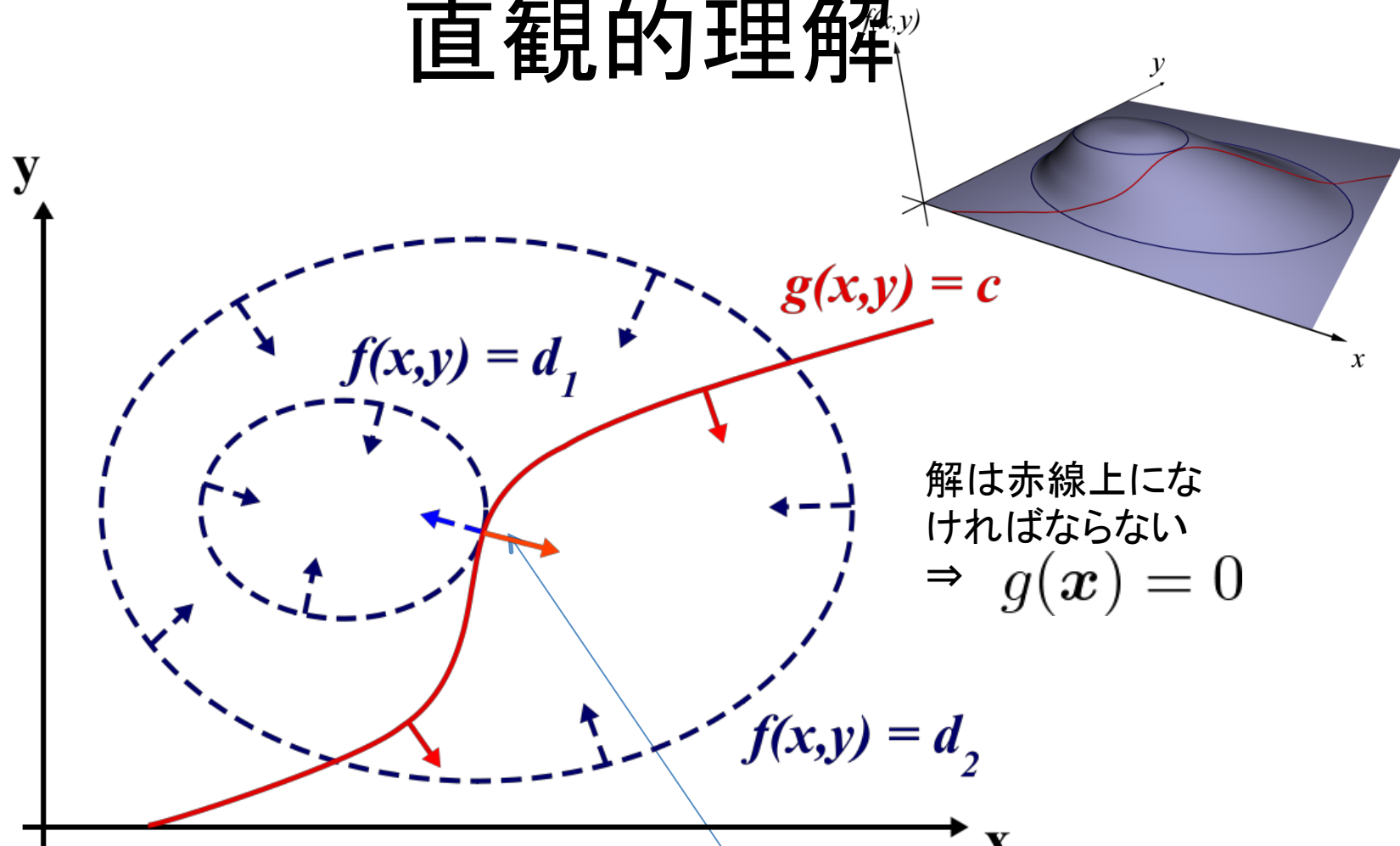
ある λ が存在して以下が成立

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) - \lambda \nabla_{\boldsymbol{x}} g(\boldsymbol{x}) = 0$$

$$\frac{\partial L(\boldsymbol{x}, \lambda)}{\partial \lambda} = g(\boldsymbol{x}) = 0$$



直観的理解



解は赤線上にな
なければならない
 $\Rightarrow g(x) = 0$

最適点 x において g と等高線は接する
 $\Rightarrow x$ における接ベクトルが平行

$$\Rightarrow \nabla_x f(x) = \lambda \nabla_x g(x)$$

なる λ が存在



分散の最大化問題へ帰着

- 問題: maximize $\mathbf{u}_1^T \Sigma \mathbf{u}_1$ subject to $\mathbf{u}_1^T \mathbf{u}_1 = 1$
- ラグランジュ緩和から求まる最適解の条件

$$\left\{ \begin{array}{l} \nabla_{\mathbf{u}_1} [\mathbf{u}_1^T \Sigma \mathbf{u}_1 - \lambda(\mathbf{u}_1^T \mathbf{u}_1 - 1)] = \mathbf{0} \\ \mathbf{u}_1^T \mathbf{u}_1 = 1 \end{array} \right.$$

このような $\mathbf{u}_1$ はどんな $\mathbf{u}_1$?



固有値問題へ帰着

- 一つめの条件

$$\nabla_{u_1} [u_1^T \Sigma u_1 - \lambda(u_1^T u_1 - 1)] = \mathbf{0}$$

$$\iff (\Sigma - \lambda I)u_1 = \mathbf{0} \quad (\text{演習})$$

- u_1 が非自明な解(非ゼロベクトル)を持つ条件
 $\det(\Sigma - \lambda I) = 0 \iff \lambda$ は Σ の固有値の一つ
- $(\Sigma - \lambda I)u_1 = \mathbf{0}$ が成り立つときの u_1
 $\Sigma u_1 = \lambda u_1 \iff u_1$ は Σ の固有ベクトルの一つ



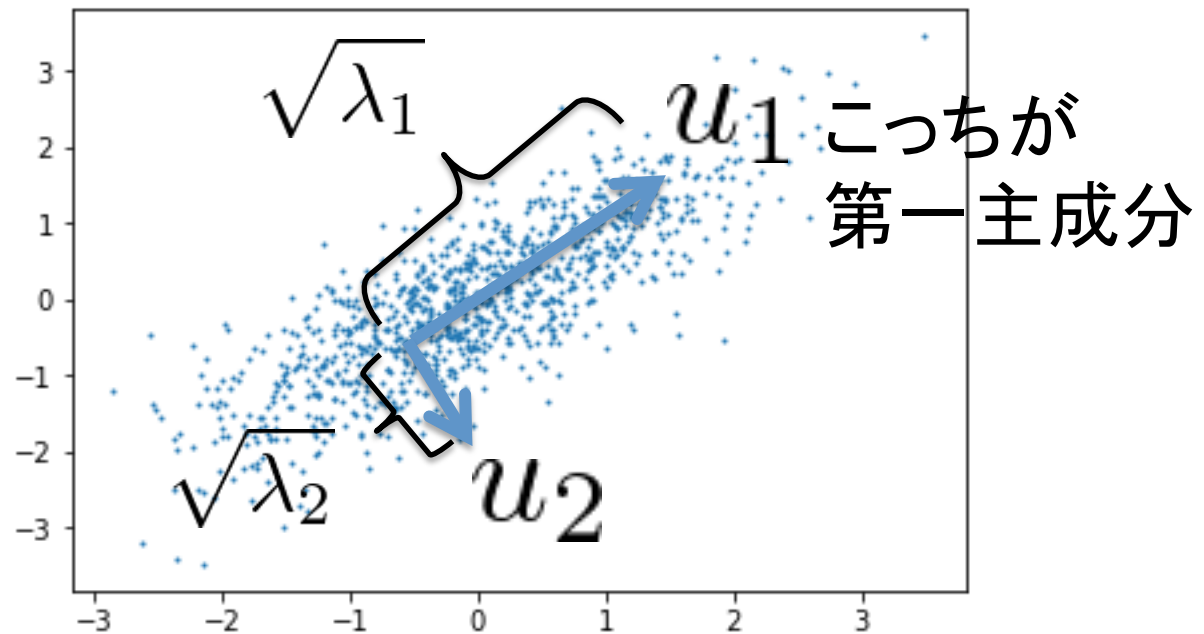
どの固有ベクトルが解なのか？

- Σ は正定値行列なのでD個の正の固有値とそれに対応する固有ベクトルを持つ
- 最大の分散 $\sigma_{z_1}^2$ を与える u_1 は結局どれ？
 - $\sigma_{z_1}^2 = \lambda$ を得る(練習)
 - 結局, 最大の分散をあたえる u_1 は、 u_1 が最大固有値に対応する固有ベクトル
 - つまり、 Σ の最大固有値に対応する固有ベクトルが、第一主成分



結論

- 第一主成分は、データの共分散行列の最大固有値を与える固有ベクトル
- 主成分問題＝固有値問題





8.6 主成分の導出 (2)

記法は 8.5 に準じる。等式制約を持つ最適化におけるラグランジュ緩和を用いれば、第一主成分方向 $\mathbf{u}_1$ は以下を満たす方向であることがわかる。

$$\nabla_{\mathbf{u}_1} [\mathbf{u}_1^T \Sigma \mathbf{u}_1 - \lambda(\mathbf{u}_1^T \mathbf{u}_1 - 1)] = \mathbf{0} \quad (10)$$

$$\mathbf{u}_1^T \mathbf{u}_1 = 1 \quad (11)$$

1. 式 41 は, $(\Sigma - \lambda \mathbf{I})\mathbf{u}_1 = \mathbf{0}$ と等価であることを示せ。
2. $\sigma_{z_1}^2 = \lambda$ であることを示せ。



第m主成分の導出

- 詳細は省略するが...
固有方程式 $(\Sigma - \lambda I)u = 0$ のm番目に大きい固有値に対応する固有ベクトルが第m主成分に対応
- Σ は共分散行列=対称半正定値行列
→固有値は全て0以上,
→固有ベクトルは正規直交基底系を構成,
→全ての主成分ベクトルは正規直交基底系を構成



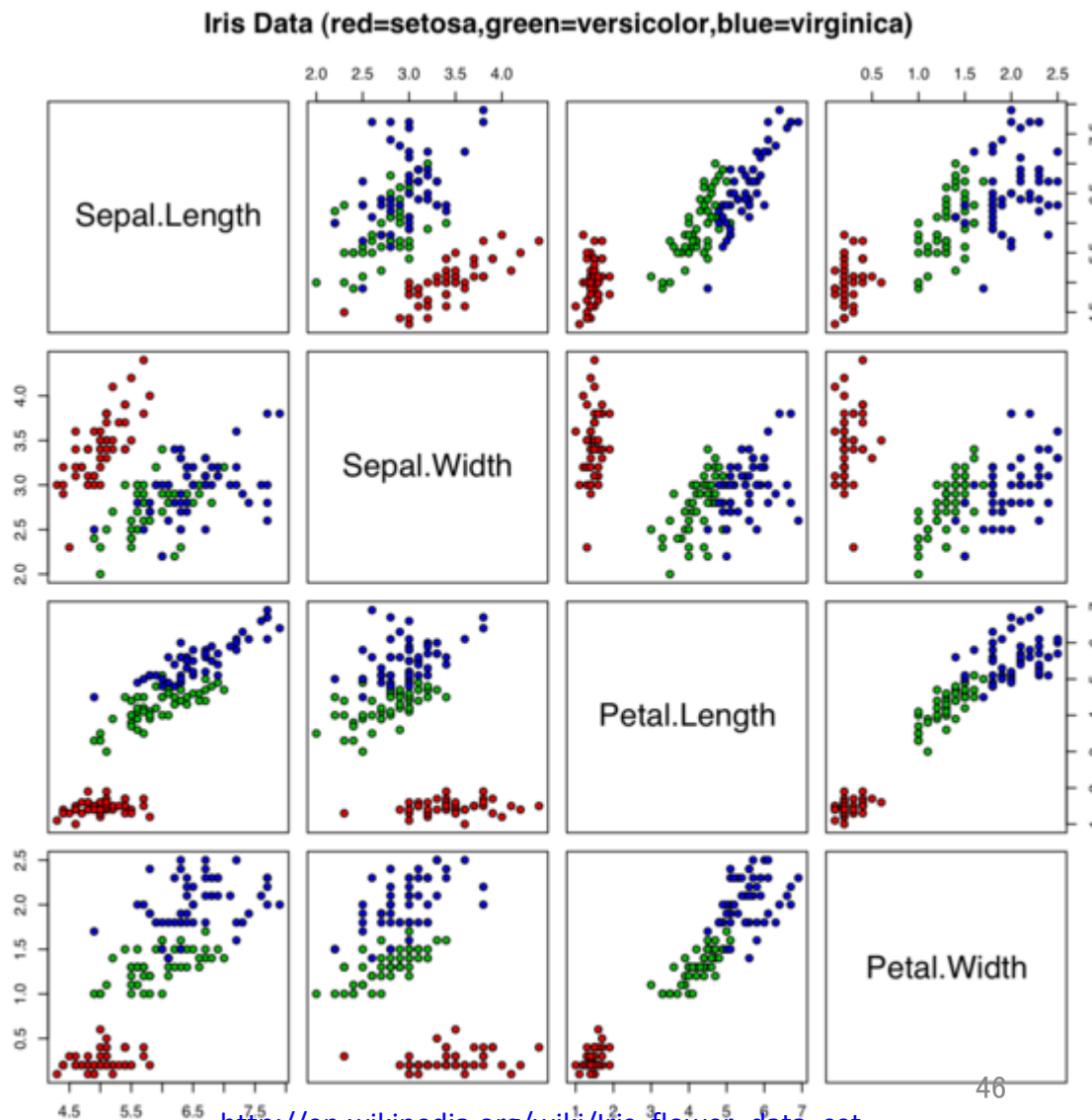
寄与率 (explained variance ratio)

- 第*i*主成分に対応する固有ベクトルの固有値を λ_i とする
- 第*i*主成分の寄与率 =
$$\frac{\lambda_i}{\sum_{d=1}^D \lambda_d}$$
- 寄与率 = 着目する主成分方向が、それ単独でどの程度サンプルが持つ情報を表現できるか
- 固有値 = その方向にデータを射影したときの分散だったことを思い出す

アヤメの特徴



- IRISデータ
- *setosa*, *versicolor*, *virginica*, 3種のあやめ
- 萼片長, 萼片幅, 花弁長, 花弁幅, 4属性値



http://en.wikipedia.org/wiki/Iris_flower_data_set



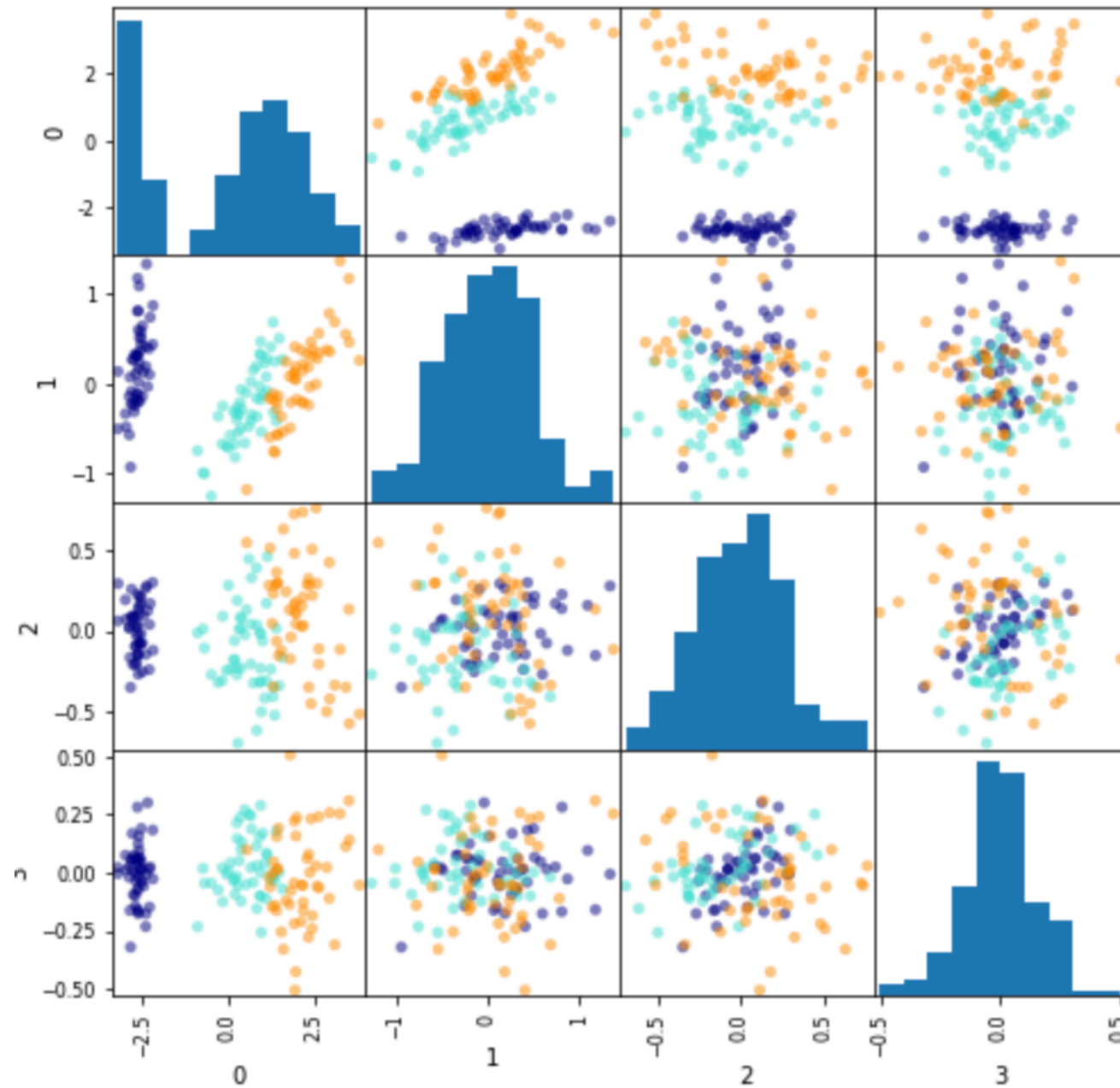
演習

8.7 主成分分析の例

プログラム PCA_iris.ipynb を実行して以下の問題に答えよ。プログラムは iris データについて主成分数 4 の主成分分析を実行し、その寄与率 (explained variance ratio) および第 i 主成分と第 j 主成分の二次元空間にデータを射影し、その散布図を表示している。上から下に $i = 1, 2, 3, 4$ の主成分, 左から右に $j = 1, 2, 3, 4$ の主成分の組み合わせで散布図が生成されている。またあやめの種類毎に異なる色の点でプロットされている。

1. 第一主成分, 第二主成分方向の組み合わせによる散布図では、3 種のアヤメは比較的よく分離して分布しているが、より番号の大きい主成分方向の組み合わせ（例えば第三主成分と第四主成分方向）による散布図では、三種のアヤメは混在して分布している。その理由を説明しなさい。
2. このアヤメが setosa かどうかを分類する分類器を構築することを考える。このとき、第一主成分の値のみを使って分類器を構築したとする。このような分類器を、4 次元の全ての特徴を用いた分類器と比較した時のメリットとデメリットを説明せよ。

PCA of IRIS dataset





PCAまとめ

- PCAは低次元かつ潜在的な構造を抽出する
 - これを, 「データの分散が最大になる方向」として定義
- ラグランジュの未定乗数法を利用した制約付き最適化に変換
- 最適解は, (中央化されたデータ行列の)固有ベクトルに対応
 - 最も分散が大きくなる方向は, 最大固有値に対応する固有ベクトル



ROADMAP

