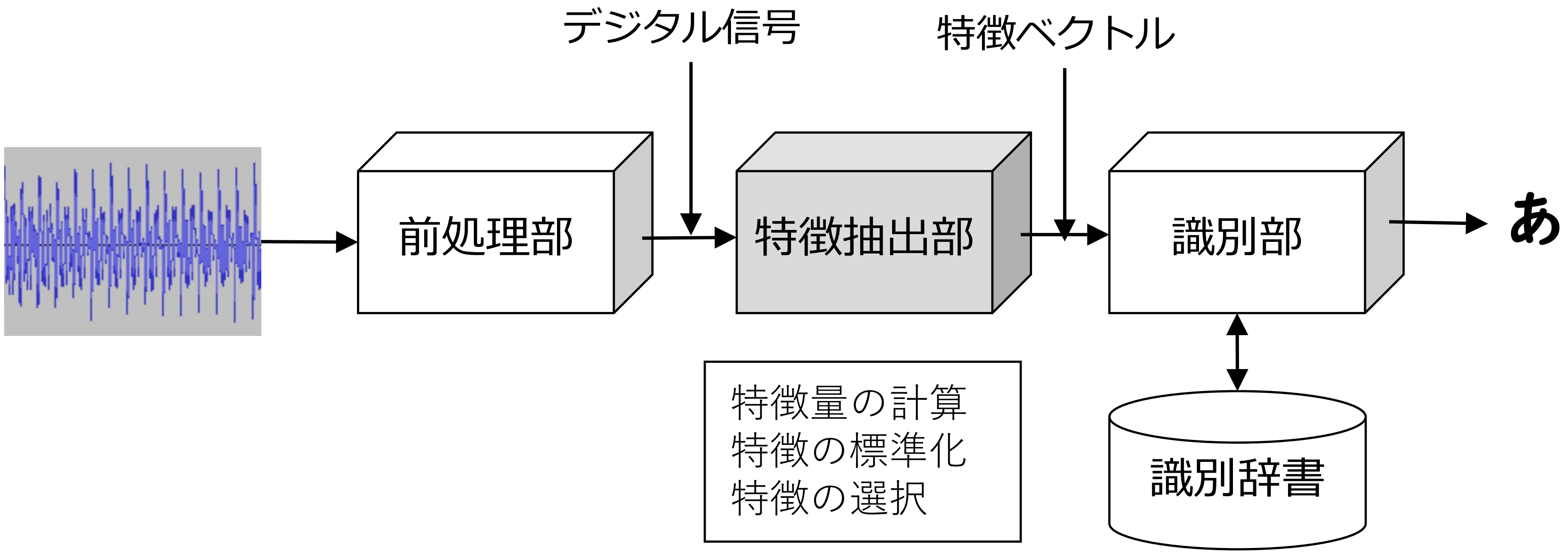
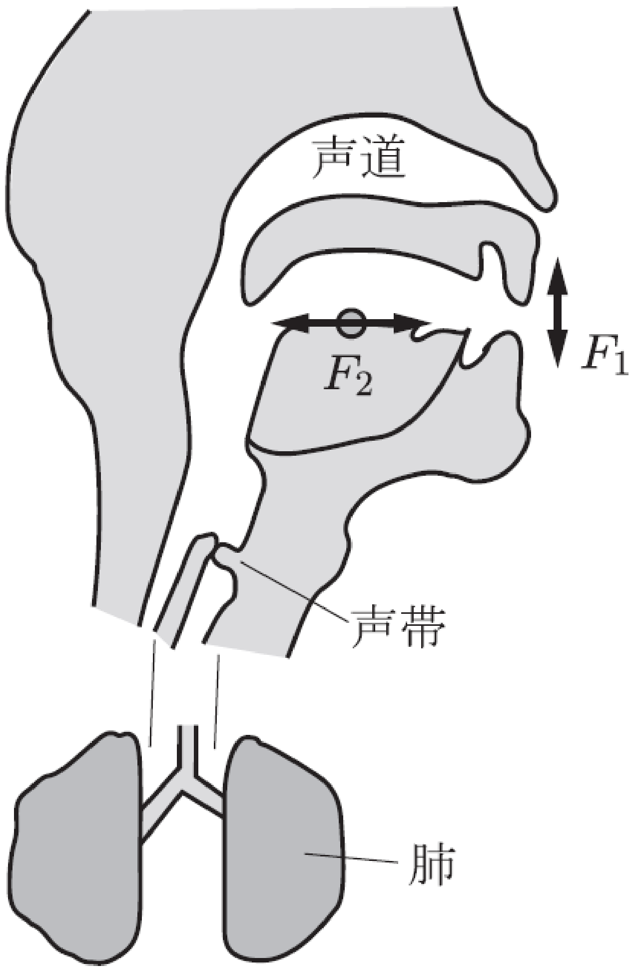
3. パターンの特徴を調べよう



3.1 変動に強い特徴とは

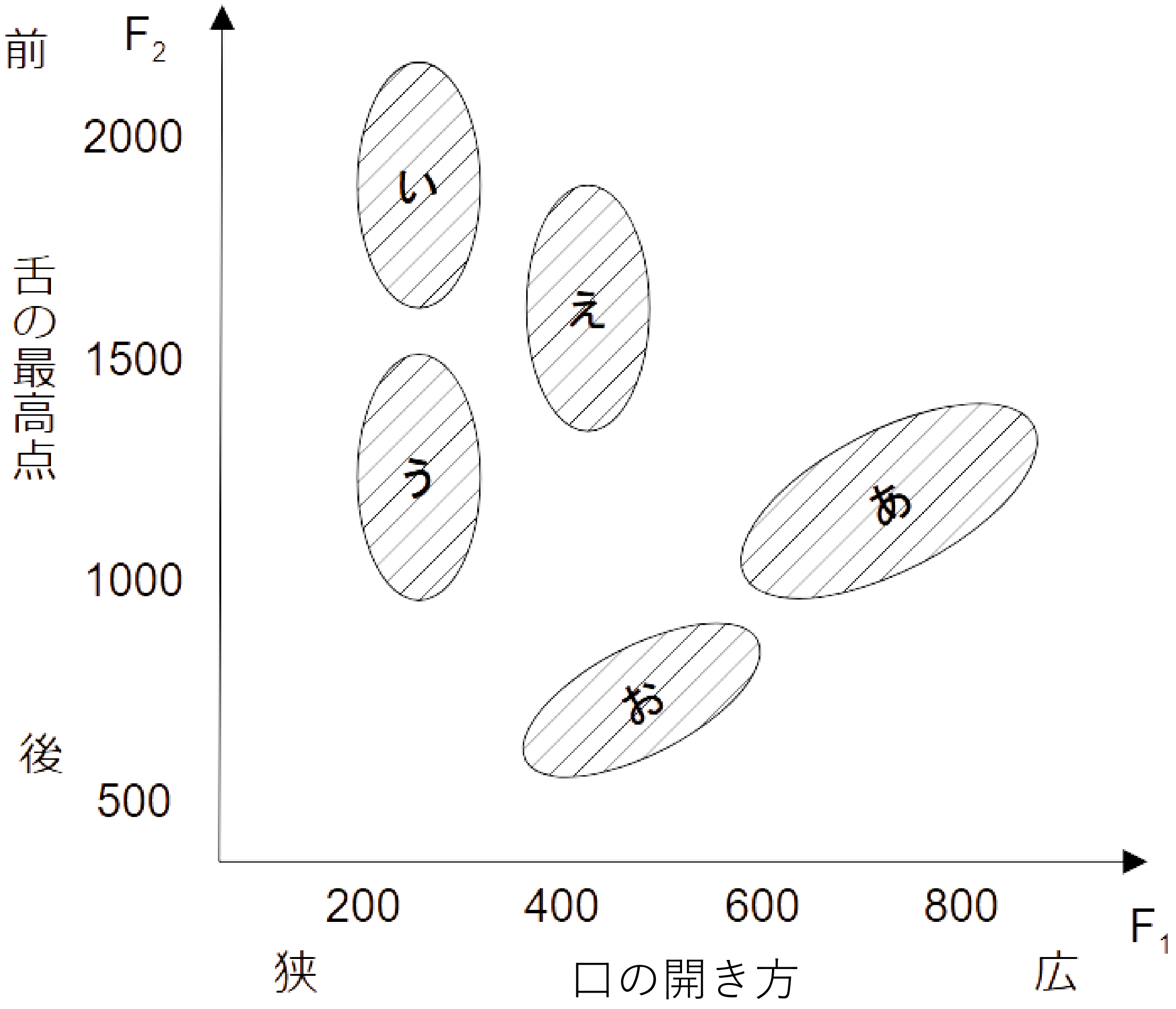
3.1.1 音声の場合

- 音素の違いとは
  - ◆ 声帯を振動の有無（パルス波か雑音か）
  - ◆ 声道（口の開き具合･舌の位置など）の変形
    →共振周波数の違いが大きな特徴



(a) 発声と調音の仕組み

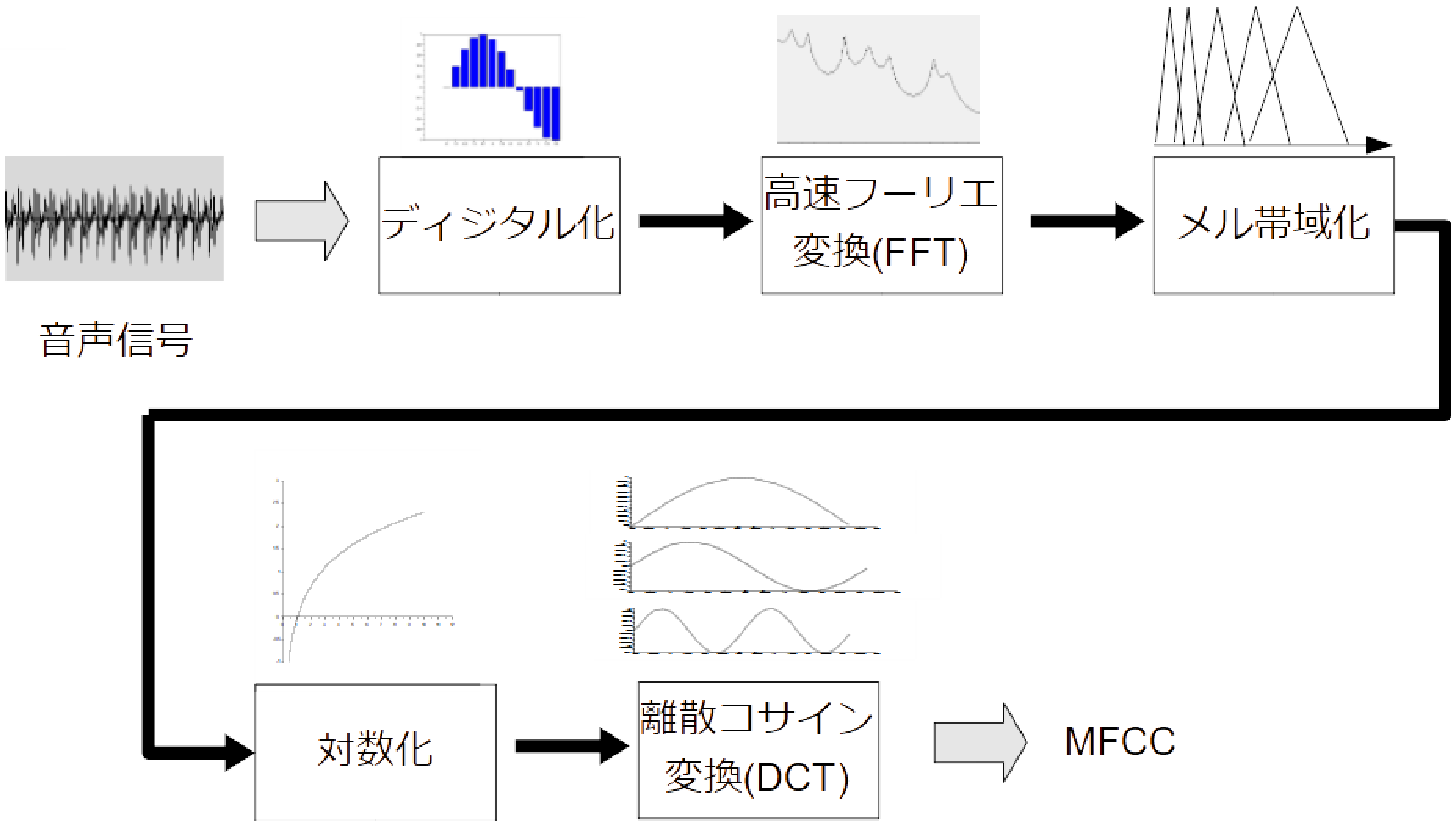
3.1.1 音声の場合



(b) 日本語母音識別のための特徴空間（男性）　軸の単位はHz

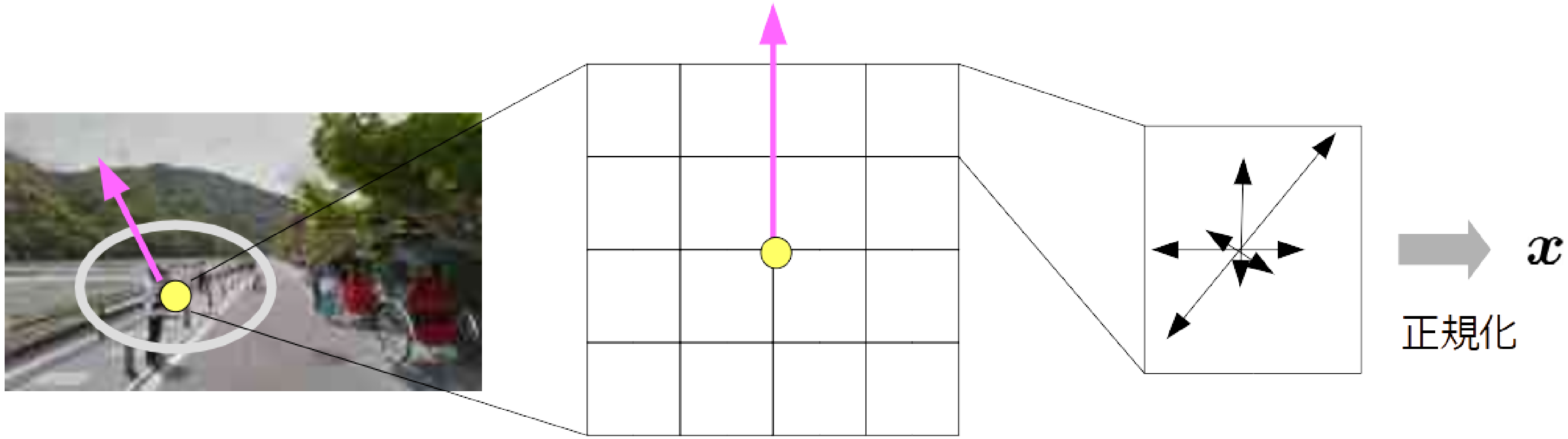
3.1.1 音声の場合

- MFCC (Mel Frequency Cepstrum Coefficient)
  - ◆スペクトルの概形情報を抽出



3.1.2 画像の場合

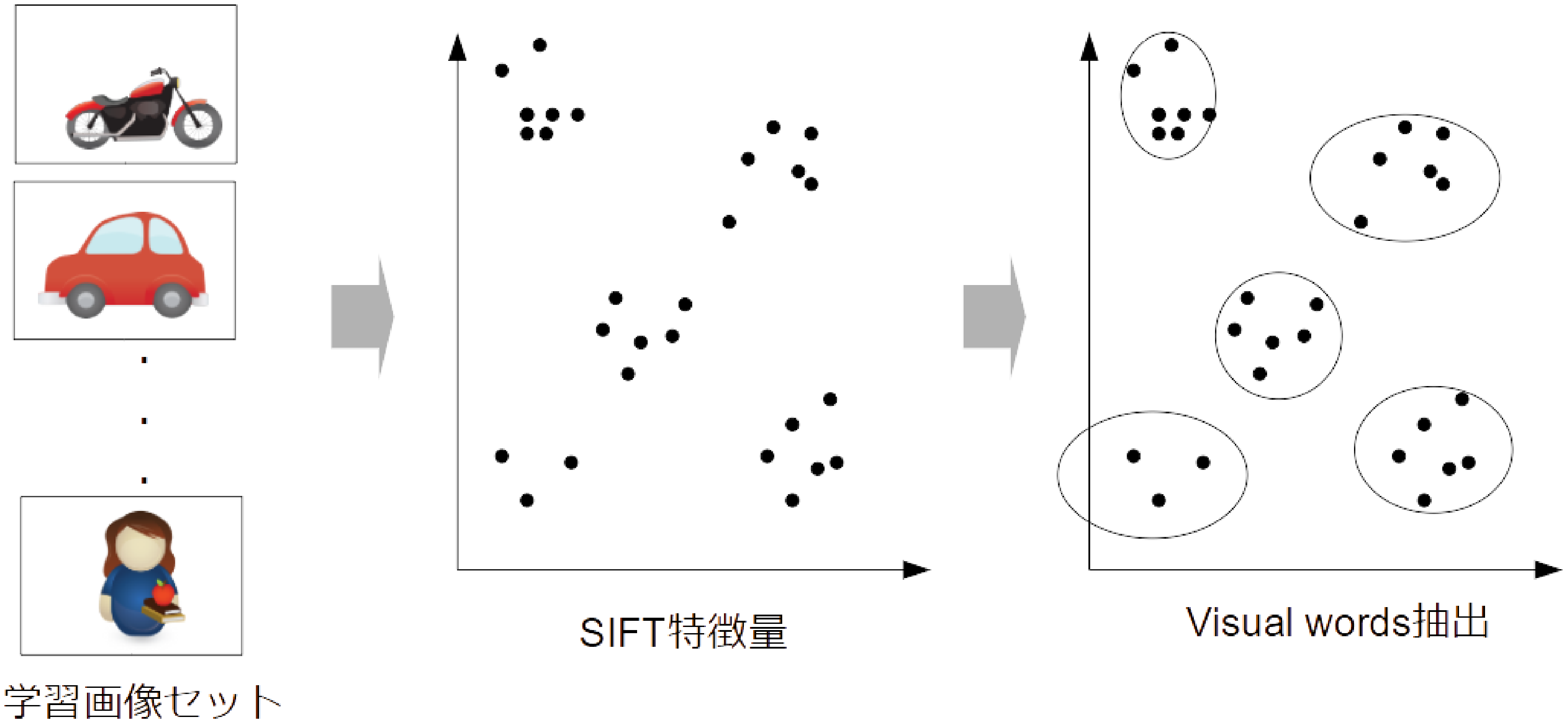
- 画像の変動
  - ◆ 明るさの変化, 拡大・縮小, 回転など
- SIFT特徴量
  - ◆ 2枚の画像の対応抽出などに有効



特徴点抽出　　スケール・回転吸収　　8方向輝度変化計算

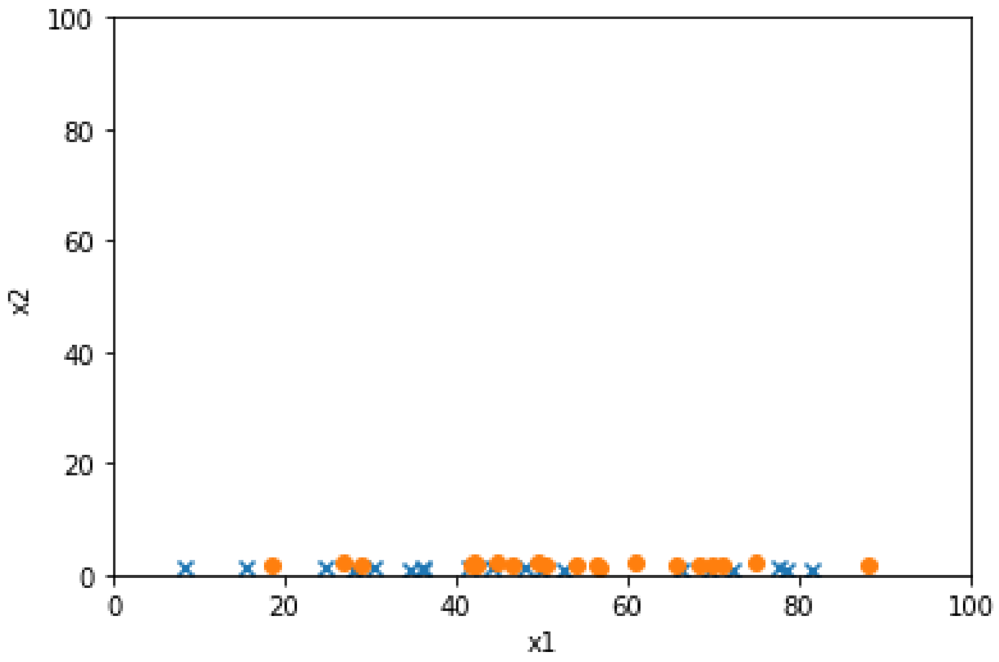
3.1.2 画像の場合

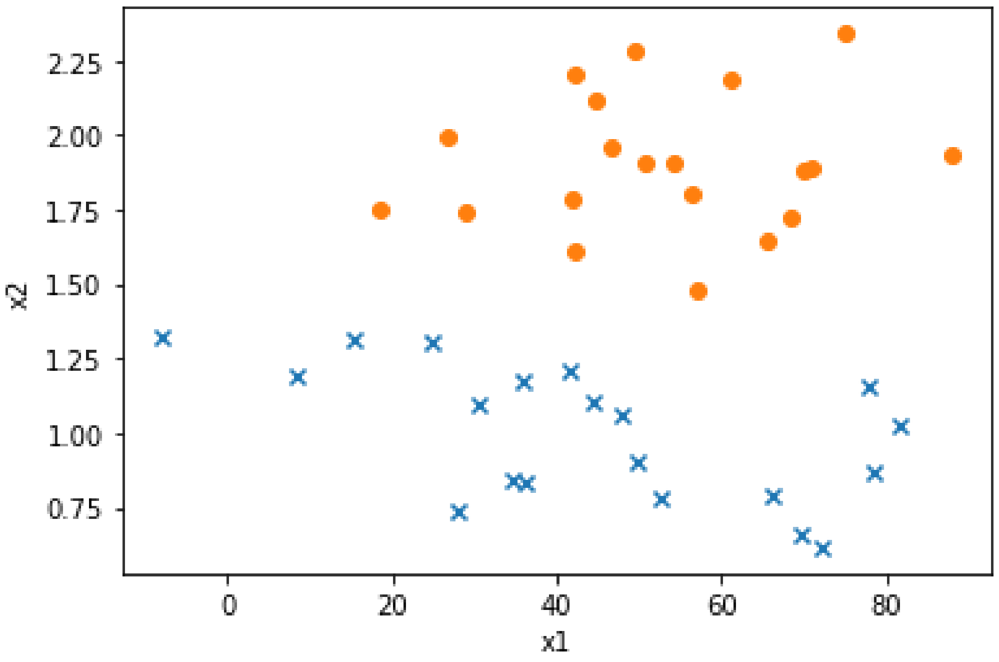
- Bag of Visual Words
  - ◆SIFT特徴量の似ているベクトルを単語と見なし、その出現頻度を特徴として識別問題に適用



3.2 特徴のスケールを揃える

- 各軸で値のスケールが異なる場合
  - ◆ 値の標準化が必要





3.2 特徴のスケールを揃える

- スケールの揃え方
  - 特徴空間の単位超立方体の体積を軸伸縮の前後で一定に保ち、かつパターン相互の距離を最小化
    →各軸の分散を等しくする
- 平均値を0にしておく
  - 学習における初期値の調整が不要
- 標準化の式

$$x'_i = \frac{x_i - m_i}{\sigma_i}$$

m_i, σ_i : 軸 i の平均、標準偏差

3.3 特徴は多いほどよいか

3.3.1 偶然に見つかってしまってはまずい

(1)偶然の傾向とは

- 特徴は多いほどよいか
  - ◆ 特徴が多くデータ数が少ないと、偶然の傾向が現れるかもしれない
  - ◆ 特徴の次元数が高いほど、偶然の傾向が発見される可能性が高い

3.3.1 偶然に見つかってしまってはまずい

(2) 学習に必要なパターン数

- 超平面の容量　$2(d+1)$
  - ◆ $p(n, d)$ ：d次元空間上で、適当に配置されたn個のパターンを任意に2クラスに分けたとき、超平面により線形分離できる確率

$$n < 2(d+1) \ : \ p(n,d) \sim 1$$
$$n = 2(d+1) \ : \ p(n,d) = 1/2$$
$$n > 2(d+1) \ : \ p(n,d) \sim 0$$

3.3.1 偶然に見つかってしまってはまずい

- 例題3.3
  - データ数：4
  - 次元数: 1

$$p(4,1)=1/2$$

○	○	○	○
○	○	○	×
○	○	×	○
○	○	×	×
○	×	○	○
○	×	○	×
○	×	×	○
○	×	×	×
×	○	○	○
×	○	○	×
×	○	×	○
×	○	×	×
×	×	○	○
×	×	○	×
×	×	×	○
×	×	×	×

(a) データの配置

○	○	○	○
○	○	○	×
○	○	×	○
○	○	×	×
○	×	○	○
○	×	○	×
○	×	×	○
○	×	×	×
×	○	○	○
×	○	○	×
×	○	×	○
×	○	×	×
×	×	○	○
×	×	○	×
×	×	×	○
×	×	×	×

(b) 二つに分離可能

3.3.1 偶然に見つかってしまってはまずい

(3) 見つかるはずのないものが見つかった？

- $n >> 2(d+1)$ のとき
  - もし、この条件で識別面が見つかったとしたら

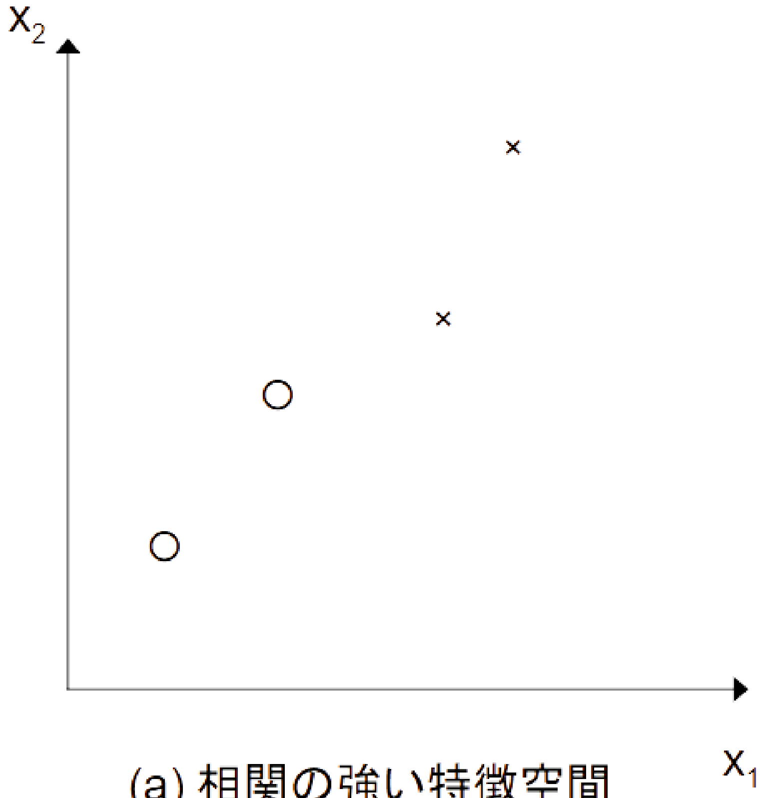
    → 偶然には存在しえないものが見つかった

    → その識別面は必然的に存在していた

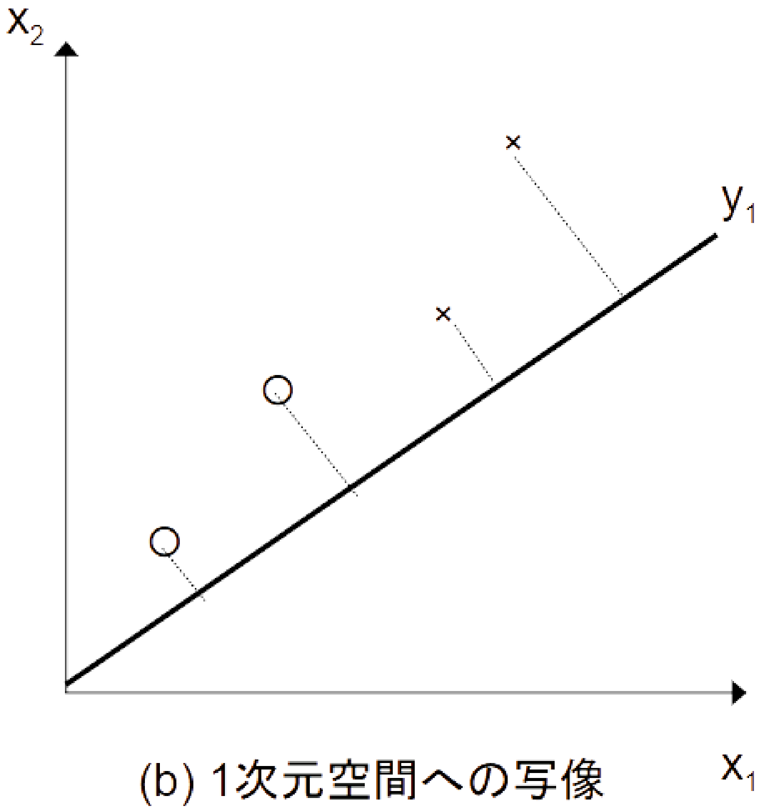
3.2.2 特徴を減らそう

(1)力業で次元を減らす → 全ての組み合わせを評価する

(2) スマートに主成分分析



(a) 相関の強い特徴空間



(b) 1次元空間への写像

共分散行列とは

- データの広がりを調べる→共分散行列
- 1次元の場合
  - ◆ 平均
  - ◆ 分散
- 多次元の場合
  - ◆ 平均ベクトル
  - ◆ 共分散行列

$$m = \frac{1}{N} \sum_{x \in \chi} x$$

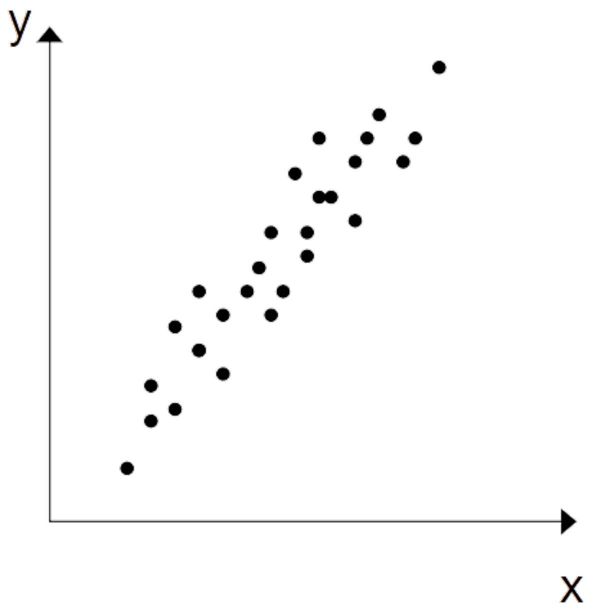
$$\sigma^2 = \frac{1}{N} \sum_{x \in \chi} (x - m)^2$$

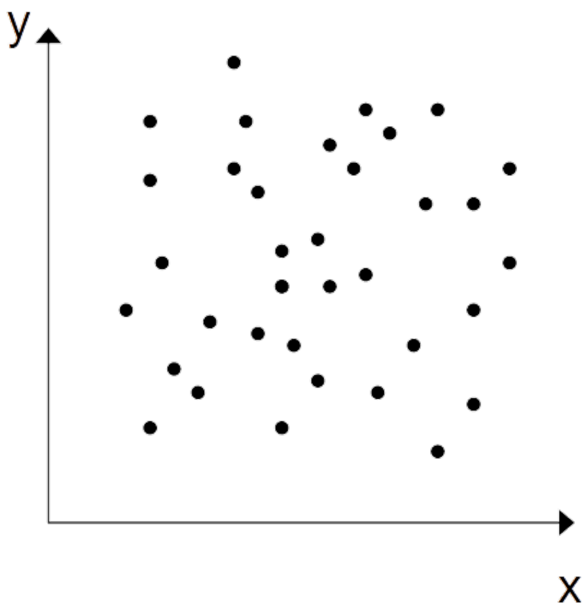
$$\boldsymbol{m} = \frac{1}{N} \sum_{\boldsymbol{x} \in \chi} \boldsymbol{x}$$

$$\boldsymbol{\Sigma} = \frac{1}{N} \sum_{\boldsymbol{x} \in \chi} (\boldsymbol{x} - \boldsymbol{m})(\boldsymbol{x} - \boldsymbol{m})^T$$

共分散行列とは

- 各軸の平均・分散が等しいデータを区別できる





$$\Sigma = \begin{pmatrix} x\text{の分散} & x\text{と}y\text{の相関} \\ x\text{と}y\text{の相関} & y\text{の分散} \end{pmatrix}$$

3.2.2 特徴を減らそう

- 主成分分析
  - ◆ 共分散行列からデータが広がっている方向（大きな固有値に対応する固有ベクトルの方向）を求める
  - ◆ そのうちのいくつかのベクトルで変換行列を構成し、高次元データを低次元データに変換する

3.2.2 特徴を減らそう

- 変換行列
  - 変換前の特徴空間におけるパターンの共分散行列Σの上位 $\tilde{d}$ 個の固有値に対応する固有ベクトルを列とする行列

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1\tilde{d}} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{d\tilde{d}} \end{pmatrix}$$

- 次元数の削減　$Y = XA$

第3章　まとめ

- 特徴抽出部の役割
  - ◆ 特徴量の計算
    - 入力の種類および認識対象によって処理が異なる
  - ◆ 特徴の標準化
    - スケールの異なる特徴の識別に対する効果を公平にする
  - ◆ 特徴の選択
    - 実験的に有効な特徴を調べる
    - 低次元に変換する