

# はじめてのパターン認識読書会

市村泰佑

May 13, 2016

## Contents

1 はじめに	1
1.1 パターン認識とは	1
1.2 特徴の型	2
1.2.1 ダミー変数	3
1.3 特徴ベクトル空間と次元の呪い	3
1.3.1 特徴ベクトル空間	3
1.3.2 次元の呪い	3
1.4 章末問題	3

## 1 はじめに

- 識別 「何か」を判断すること
- 特徴抽出 「有効な特徴」を抽出すること
- 識別規則 「有効な特徴」と「何か」を結びつける規則

### 1章で学ぶこと

- パターン認識の流れ
- パターン認識に用いられる特徴の種類とその性質

### 1.1 パターン認識とは

辞書では

- 「パターン」 型、類型
- 「認識」 人間が物事を知る働き、およびその内容

身の回りにおけるパターン認識装置  
駅の切符券売機における硬貨や、紙幣の種類識別

1. 硬貨や紙幣の材質やサイズなど識別の手がかりとなる特徴を測定すること「特徴抽出」
2. 硬貨の重さ、サイズ、穴の有無などの特徴を並べてベクトルの形にしたもの「特徴ベクトル」
3. 特徴ベクトルを用いて入力硬貨がどの硬貨なのかクラスに分類する、この分類するための規則のこと「識別規則」

識別に有効な特徴をいかに早く抽出できるかがパターン認識の成否を決める鍵

パターン認識の対象は多岐にわたり、特徴抽出の方法も多岐にわたるが、一旦特徴抽出され、特徴ベクトルの形になれば、観測対象が多岐にわたっても同じ識別規則を使うことができる。

特徴ベクトルを入力データとし、クラス分けを行うための識別規則とその学習法について学ぶ

学習データ … 識別規則を作るためには、入力データとそのクラスを対にしたたくさんの事例を使って入力データとクラスの対応関係を学習する必要がある。この事例のこと

識別規則の学習がパターン認識の成否を決めるもう一つの鍵でこの本の主題

## 1.2 特徴の型

パターン認識は特徴抽出から始まる。観測される特徴には二つに大別される。

- 定性的特徴 非数値データとして抽出される特徴
- 定量的特徴 数値データとして抽出される特徴

非数値データは名義尺度、順序尺度に細分される

尺度名	定義	例
名義尺度	分類のための名前, クラスラベル	名前, 住所, 電話番号など
順序尺度	順序関係	大中小, 優良可
間隔尺度	一定の単位で測られた量で、等間隔性がある	試験成績, 年月
比例尺度	原点が定まっている量, 比が意味を持つ	身長, 体重

### 1.2.1 ダミー変数

クラス数が  $K$  個の場合、 $K$  個の 2 値変数を用意し、クラスに対応する変数のみを 1, その他を 0 というように符号化をし、クラスレベルを表す。

	$\beta_1$	$\beta_2$	$\beta_3$
大	1	0	0
中	0	1	0
小	0	0	1

## 1.3 特徴ベクトル空間と次元の呪い

### 1.3.1 特徴ベクトル空間

特徴数が  $d$  とすれば、特徴ベクトルは  $d$  次元線形空間を張る。 $16 \times 16 = 256$  の画素を持つ画像データでは 256 次元ベクトル空間中の 1 点として表される。

### 1.3.2 次元の呪い

前述の各画素が 16 レベルの階調を持つとすると 256 次元のベクトル空間の各軸が 16 個の区画を持つこととなり  $16^{256}$  と、次元に対して指数関数的に増加する。このことを 次元の呪い と呼ぶ

1.  $d$  次元単位超立方体の中心から頂点までの距離

$$D(d) = (d(\frac{1}{2})^2)^{\frac{1}{2}} = \frac{1}{2}\sqrt{d}$$

中心から単位超立方体の面までの最短距離は面によらず  $\frac{1}{2}$  であるため頂点までの距離との比は  $\sqrt{d}$  となる。

## 1.4 章末問題

1. あなたの利き手でない方の人差し指と中指の指紋を区別したい。どのような特徴を取れば良いか観察せよ
2. 辺の長さが  $a$  の  $d$  次元立方体について、以下の問いに答えよ
  - (a)  $d$  次元超立方体の頂点の数は  $2^d$  個あることを示せ  
1 次元空間では直線となるため、頂点は 2 点。2 次元空間では平面となるため 4 点。3 次元空間では 8 点となる。同様に  $d$  次元空間では  $2^d$  個の頂点を持つ。

(b)  $d$ 次元超立方体の表面積を求めよ

辺の長さを  $a$  とすると二次元空間では  $a^2$ 。三次元空間では  $a^2 * 6$ 。  
つまり、 $a^{d-1} * 2d$

(c) 超立方体を構成する  $m$  次元超平面 ( $0 \leq m \leq d - 1$ ) の個数が

$2^{d-m} \binom{d}{m}$  で表されることを、3次元立方体で確かめよ。

面である二次元超平面の個数は

$$m = 2$$

$$d = 3$$

$$2^{3-2} \binom{3}{2}$$

線である1次元超平面の個数は

$$m = 1$$

$$d = 3$$

$$2^{3-1} \binom{3}{1}$$

点である0次元超平面の個数は

$$m = 0$$

$$d = 3$$

$$2^{3-0} \binom{3}{0}$$

(d) 超立方体を構成する  $m$  次元超平面の総数を求めよ

$$\sum_{m=0}^{d-1} 2^{d-m} \binom{d}{m}$$

(e) その式から、5次元超立方体を構成する超平面の総数を求めよ