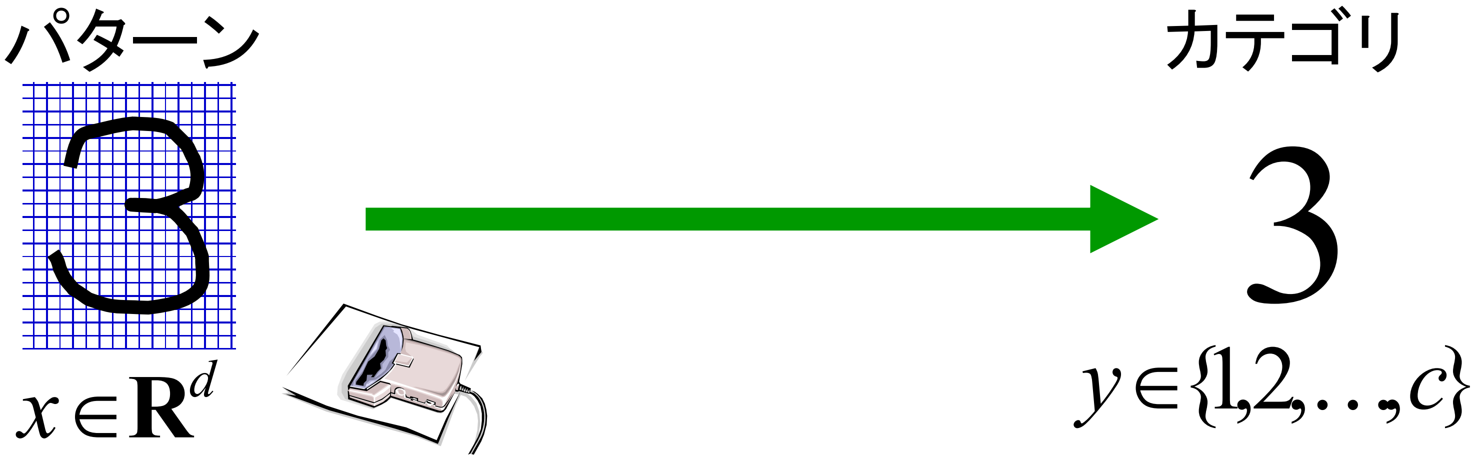
統計的機械学習
第1回　確率統計の基礎

2013/05/28
表現工学科　尾形研究室　ゼミ
野田　邦昭

パターン認識

- 入力パターンをカテゴリに割り当てる



確率変数

- カテゴリ y やパターン x を確率変数(random variable)として扱えば，次のような「確率」が定義できる．

$$p(x), p(y), p(x, y), p(y \mid x), p(x \mid y)$$

- カテゴリ y : 離散型(discrete type)の確率変数
- パターン x : 連続型(continuous type)の確率変数

確率関数と確率密度関数

- $p(y)$:カテゴリ y の生起確率を表す確率関数(probability function)

$$\sum_{y=1}^{m} p(y) = 1$$

$$p(y) \geq 0 \ \text{ for } \ y = 1,2,\ldots,m$$

- $p(x)$:パターン x の確率密度関数(probability density function)

$$\int_D p(x)dx = 1$$

$$p(x) \geq 0 \ \text{ for all } x \in D$$

同時確率と条件付き確率

- $p(x,y)$: x と y の同時確率(joint probability)
- 周辺化(marginalization):

$$\sum_{y=1}^{m} p(x,y) = p(x)$$

$$\int_D p(x,y)dx = \underbrace{p(y)}_{\text{周辺確率 (marginal probability)}}$$

- $p(x|y), p(y|x)$: 条件付き確率(conditional probability)

$$p(y|x)p(x) = p(x,y) = p(x|y)p(y)$$

事前確率・事後確率・ベイズの定理

- 事前確率(a priori probability) $p(y)$ ：
  パターンを知る前のカテゴリの出現確率
- 事後確率(a posteriori probability) $p(y \mid x)$ ：
  パターンを知った後のカテゴリの出現確率
- ベイズの定理(Bayes' theorem):

$$p(y \mid x) = \frac{p(x \mid y)\,p(y)}{p(x)}$$

期待値と分散共分散

- 期待値 (Expectation)

$$\boxed{\mathrm{E}_{\mathbf{x}}[\mathbf{x}] := \int \mathbf{x} p(\mathbf{x}) d\mathbf{x}} \qquad \mathrm{E}_{\mathbf{x}}\left[f(\mathbf{x})\right] := \int f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

- 分散共分散 (Variance-covariance)

$$\boxed{\mathrm{V}_{\mathbf{x}}[\mathbf{x}] := \mathrm{E}_{\mathbf{x}}\left[\left(\mathbf{x} - \mathrm{E}_{\mathbf{x}}[\mathbf{x}]\right)\left(\mathbf{x} - \mathrm{E}_{\mathbf{x}}[\mathbf{x}]\right)^T\right]}$$

分散 $\left[\mathrm{V}_{\mathbf{x}}[\mathbf{x}]\right]_{i,i} = \mathrm{V}_{\mathbf{x}}[\mathbf{x}^{(i)}] = \mathrm{E}_{\mathbf{x}}\left[\left(\mathbf{x}^{(i)} - \mathrm{E}_{\mathbf{x}}[\mathbf{x}^{(i)}]\right)^2\right]$

共分散 $\left[\mathrm{V}_{\mathbf{x}}[\mathbf{x}]\right]_{i,j} = \mathrm{C}_{\mathbf{x}}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)$

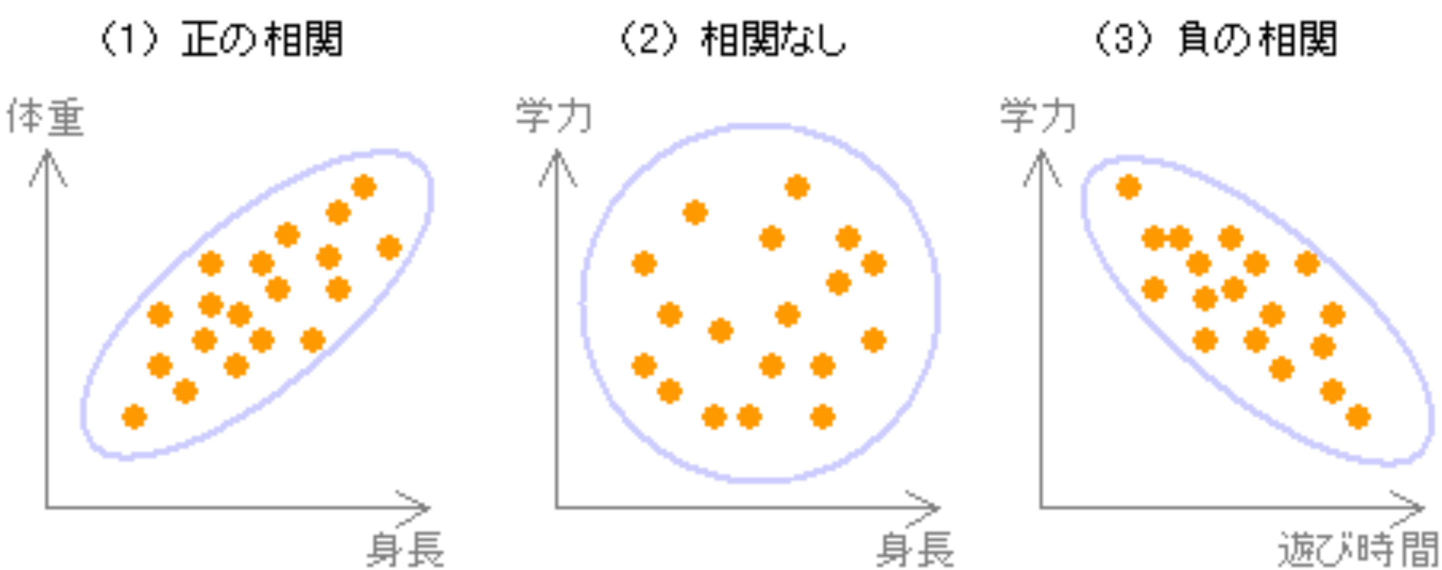
$$:= \mathrm{E}_{\mathbf{x}}\left[\left(\mathbf{x}^{(i)} - \mathrm{E}_{\mathbf{x}}[\mathbf{x}^{(i)}]\right)\left(\mathbf{x}^{(j)} - \mathrm{E}_{\mathbf{x}}[\mathbf{x}^{(j)}]\right)\right]$$

※期待値はd次元ベクトル，分散共分散はdxd行列

相関係数 (Correlation coefficient)

- 共分散を正規化したもの

$$\rho = \frac{\mathrm{C}_{\mathbf{x}}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)}{\sqrt{\mathrm{V}_{\mathbf{x}}[\mathbf{x}^{(i)}]}\sqrt{\mathrm{V}_{\mathbf{x}}[\mathbf{x}^{(j)}]}}$$



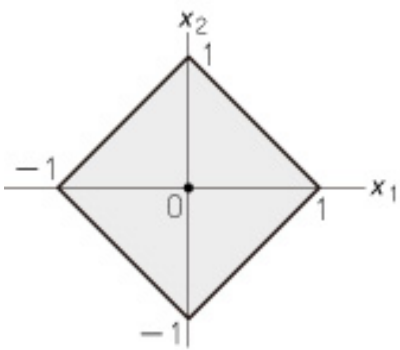
無相関と独立の違い

- 無相関とは
  - 相関係数が限りなく0に近いこと

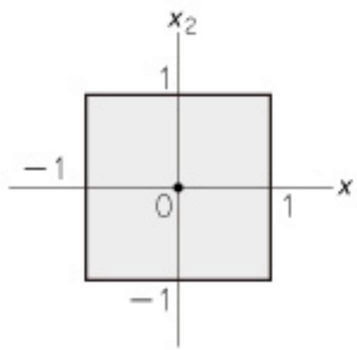
$$\rho = 0$$

- 独立とは
  - 同時確率が下のように分解できること

$$p(\mathbf{x}, \mathbf{x}') = p(\mathbf{x})p(\mathbf{x}')$$



(a) 無相関



(b) 独立

$$p(x, x') = \begin{cases} 1 & |x| + |x'| \leq 1 \\ 0 & \text{else} \end{cases}$$

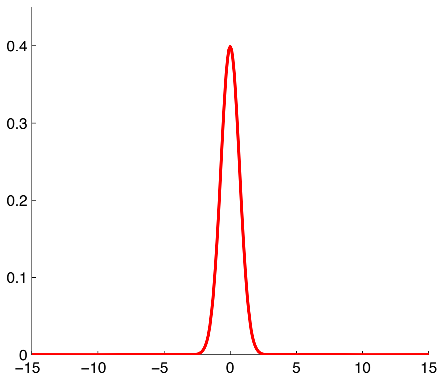
$$p(x, x') = \begin{cases} 1 & |x| \leq 1, |x'| \leq 1 \\ 0 & \text{else} \end{cases}$$

独立ならば，無相関だが，逆は必ずしも成り立たない

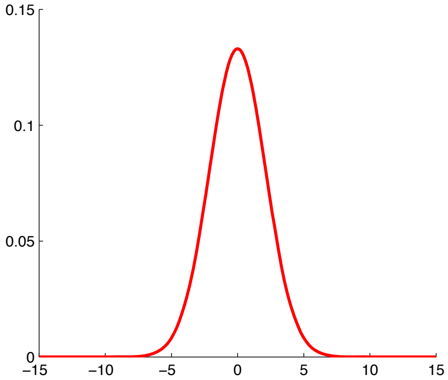
正規分布

- ２つのパラメータ: μ, σ^2

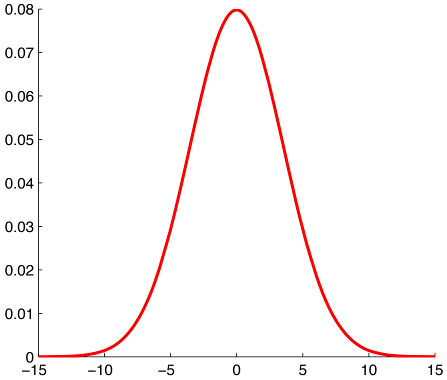
$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$



$\sigma^2 = 1$



$\sigma^2 = 9$



$\sigma^2 = 25$

- 正規分布の平均と分散:

$$E[x] = \mu \qquad V[x] = \sigma^2$$

多次元正規分布

- d次元の確率ベクトル: $x = (x^{(1)}, x^{(2)}, \ldots, x^{(d)})^T$
- 2つのパラメータ:
  - d 次元ベクトル μ
  - d 次元正値行列 Σ

$\det(\Sigma)$: Σの行列式

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

- 正規分布の期待値, 分散共分散行列

$$E[x] = \mu \qquad V[x] = \Sigma$$

多次元正規分布(つづき)

■ 共分散がゼロ(即ち $\Sigma = diag(\sigma_i^2)$)のとき

$diag(\sigma_i^2)$: $\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2$ を対角成分に持つ対角行列

$$p(x; \mu, \{\sigma_i^2\}_{i=1}^d) = \frac{1}{(2\pi)^{d/2} \prod_{i=1}^d \sigma_i} \exp\left(-\sum_{i=1}^d \frac{(x^{(i)} - \mu^{(i)})^2}{2\sigma_i^2} \right)$$

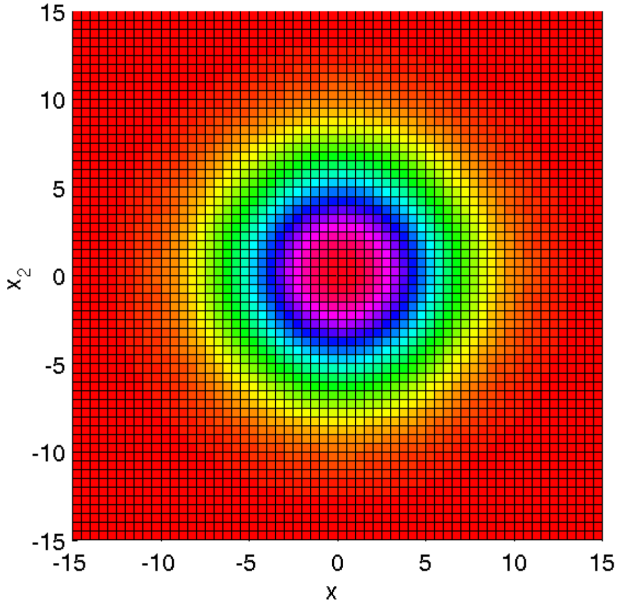
■ さらに分散が等しい(即ち $\Sigma = \sigma^2 I$)のとき

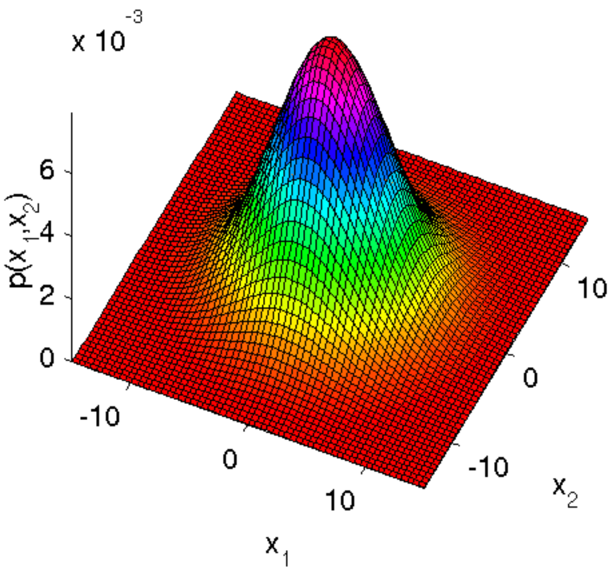
I :単位行列

$$p(x; \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(x-\mu)^T (x-\mu)}{2\sigma^2} \right)$$

多次元正規分布の例(1)

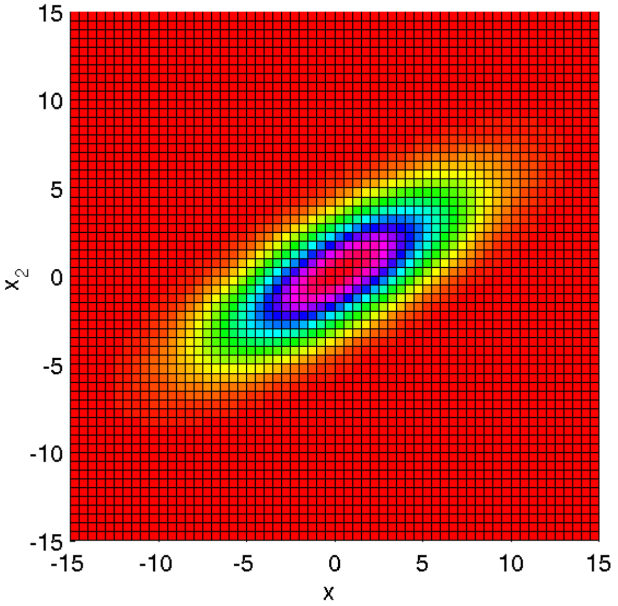
$d = 2$ $\mu = (0,0)^T$ $\Sigma = \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix}$

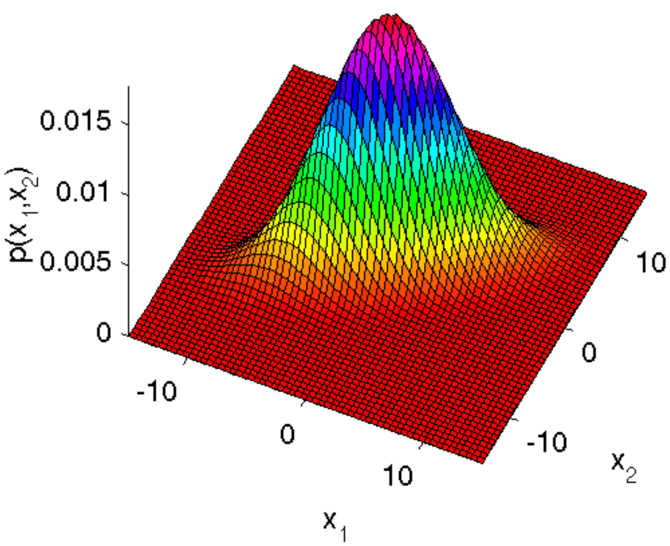




多次元正規分布の例(2)

$d = 2$ $\mu = (0,0)^T$ $\Sigma = \begin{pmatrix} 20 & 10 \\ 10 & 9 \end{pmatrix}$



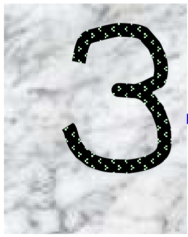


パターン認識とは？

- パターン認識とは:「与えられたパターンをそれが属するカテゴリに対応づける操作」
- パターンとは:「空間的や時間的に観測可能な事象であって，観測された事象どうしが同一であるか否かを判定できるような性質を備えているもの」
- カテゴリとは:「パターン認識の結果，同等とみなされるパターンの集合概念」

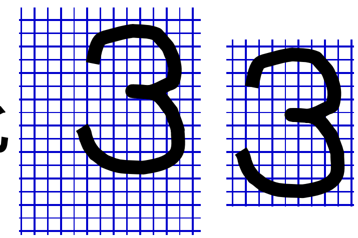
パターン認識の過程

入力パターン

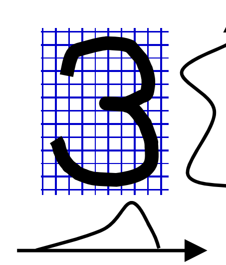


観測 → 入力パターンの取り込み

前処理 → 雑音除去や正規化



特徴抽出 → 認識に有益な情報の抽出



識別 → カテゴリの決定

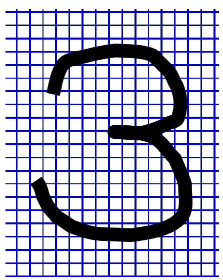
出力カテゴリ

3

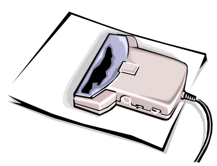
パターン認識

- 入力パターンをカテゴリに割り当てる

パターン



$x \in \mathbf{R}^d$



カテゴリ

3

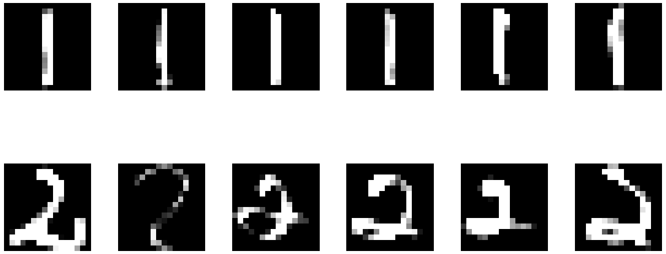
$y \in \{1, 2, \ldots, c\}$

パターンとカテゴリの表記

- パターン(pattern) x : d 次元実ベクトル
- パターン空間(pattern space) $D\left(\subset \Re^{d}\right)$:
  パターンの定義域(domain)
- y :カテゴリ(category) $y \in \{1,2,\ldots,c\}$
- c :カテゴリの数

手書き文字認識の例

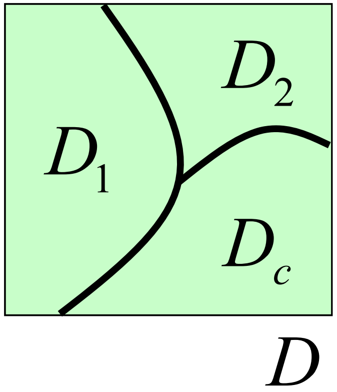
- スキャナで取り込んだ文字画像が16×16画素のとき，パターン x は各画素の濃度を縦に並べた256次元のベクトル．
- 厳密には画素値は実数ではない（例えば8ビット，即ち256階調の離散値）が，$[0,1]$ に正規化した実数値として扱う．
- このとき，パターン空間は $D=[0,1]^{256}$ ．
- カテゴリは各文字に対応．



識別関数・決定領域・決定境界

- 識別関数(discrimination function) $f(x)$: パターン x をそれが属するカテゴリ y に対応づける関数
- 決定領域(decision region) D_y : カテゴリ y のパターンが属する領域
- 決定境界(decision boundary) : いくつかの決定領域どうしの境界

識別関数を求めること
＝決定領域を求めること
＝決定境界を求めること



統計的パターン認識

- 識別関数(決定領域，決定境界)は未知

- 統計的パターン認識 (statistical pattern recognition):カテゴリ y やパターン x を確率変数として扱い，それらの統計的な性質を利用してパターン認識を行う

教科書

■ 統計的機械学習，杉山将著，オーム社

参考書

- パターン認識と機械学習：ベイズ理論による統計的予測（上・下），C. M. Bishop著，元田浩 他（訳）
- 認識工学：パターン認識とその応用，鳥脇純一郎著（コロナ社）
- わかりやすいパターン認識，石井健一郎 [ほか] 著，オーム社
- パターン識別，R.O.Duda, P.E.Hart, D.G.Stork著（尾上守夫監訳），新技術コミュニケーションズ
- パターン認識とニューラルネットワーク，栗田多喜夫著 (http://www.neurosci.aist.go.jp/~kurita/lecture/prnn.pdf)
- 統計学入門，東京大学教養学部統計学教室編，東京大学出版会