

# 知能機械と自然言語処理

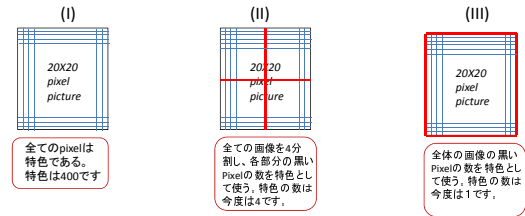
## 知能機械部 第3回

ソフトウェア情報学部

ゴウタム

1

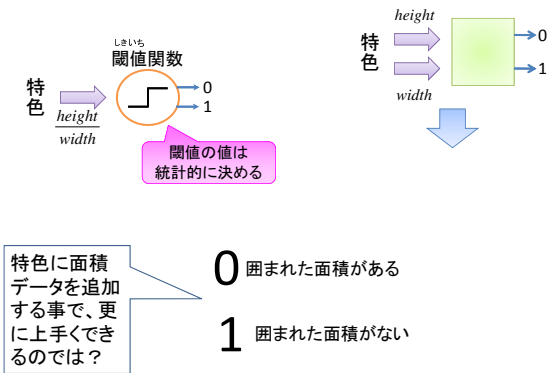
# 特色について



(I), (II), (III)特色のどれを使用することで認識率を一番高める事が出来ますか？  
実験すると(II)が一番高いです。

2

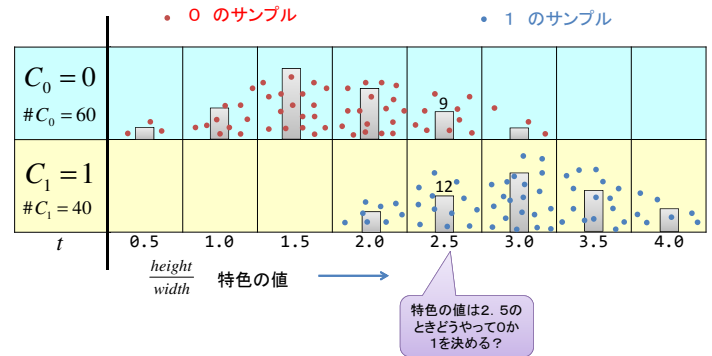
# 前回の課題の補足



0 囲まれた面積がある  
1 囲まれた面積がない

3

# 統計学の復習



4

# 統計用語 ~ 確率

離散確率: サンプルは数えるとき      連続確率: サンプルは計るとき

$i$  はClass番号、 $(C_0)$  はClass 0 です  $t$  は特色の値です

## Joint Probability (結合確率)

$$P(C_i, t) \quad i=0, t=2.5 \text{ のとき、} \\ P(C_0, t=2.5) = \frac{n(C_0 \cap t)}{n(U)} = \frac{9}{100}$$

## Class Conditional Probability (Class条件付確率)

$$P(t | C_i) \quad i=0, t=2.5 \text{ のとき、} \\ P(t=2.5 | C_0) = \frac{n(t \cap C_0)}{n(C_0)} = \frac{9}{60}$$

Class Conditional Probability

5

# ベイズの定理【公式】

$$P(C_i, t) = P(t | C_i) P(C_i) = P(C_i | t) P(t)$$

$$\therefore P(C_i | t) = \frac{P(t | C_i) P(C_i)}{P(t)}$$

Posterior Probability (事後確率)  $P(t)$  ← 全てのClassが一定なので計算不要

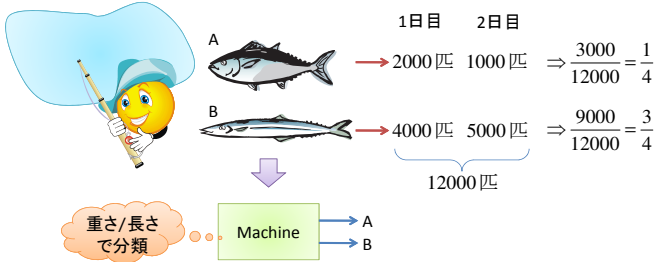
Class 0 と Class 1 のデータからClassの条件付き確率  $P(t | C_0), P(t | C_1)$  を求める。 $P(C_i)$  は予め分かる。そうすることで事後確率を求められます。特色  $t$  のサンプルは事後確率が高いほうのClassだと判断することが出来る

6

## 統計用語 ~ 基礎

### • A Prior Probability (事前確率)

$$P(C_0) = \frac{60}{100} = \frac{3}{5}, \quad P(C_1) = \frac{40}{100} = \frac{2}{5} \quad \therefore \sum_k^n P(C_k) = 1$$



7

## ベイズの定理【例】

$$P(C_0 | 2.5) = \frac{P(2.5 | C_0)P(C_0)}{P(2.5)} = \frac{\frac{9}{21} \times \frac{60}{100}}{\frac{9}{21}} = \frac{9}{21}$$

$$P(C_1 | 2.5) = \frac{P(2.5 | C_1)P(C_1)}{P(2.5)} = \frac{\frac{12}{40} \times \frac{40}{100}}{\frac{12}{21}} = \frac{12}{21}$$

よって、 $t = 2.5$  は Class  $C_1$  と判断する。

$$\therefore \sum_k^n P(C_k | t) = 1$$

8

## 統計的(分類)手法

- Parametric
  - BAYES
- Non- Parametric
  - Parzen Window (パルザン窓)
  - N-Nearest Neighbor (N近傍)

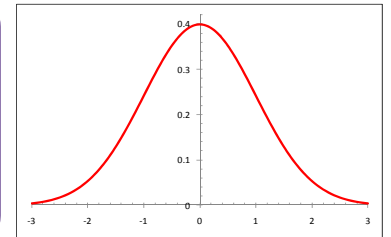
9

## 正規分布

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

$$\text{または、} f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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



- 自然のデータのほとんどが正規分布を成している

10