

# 機械学習

## 統計的機械学習（生成モデル）

管理工学科

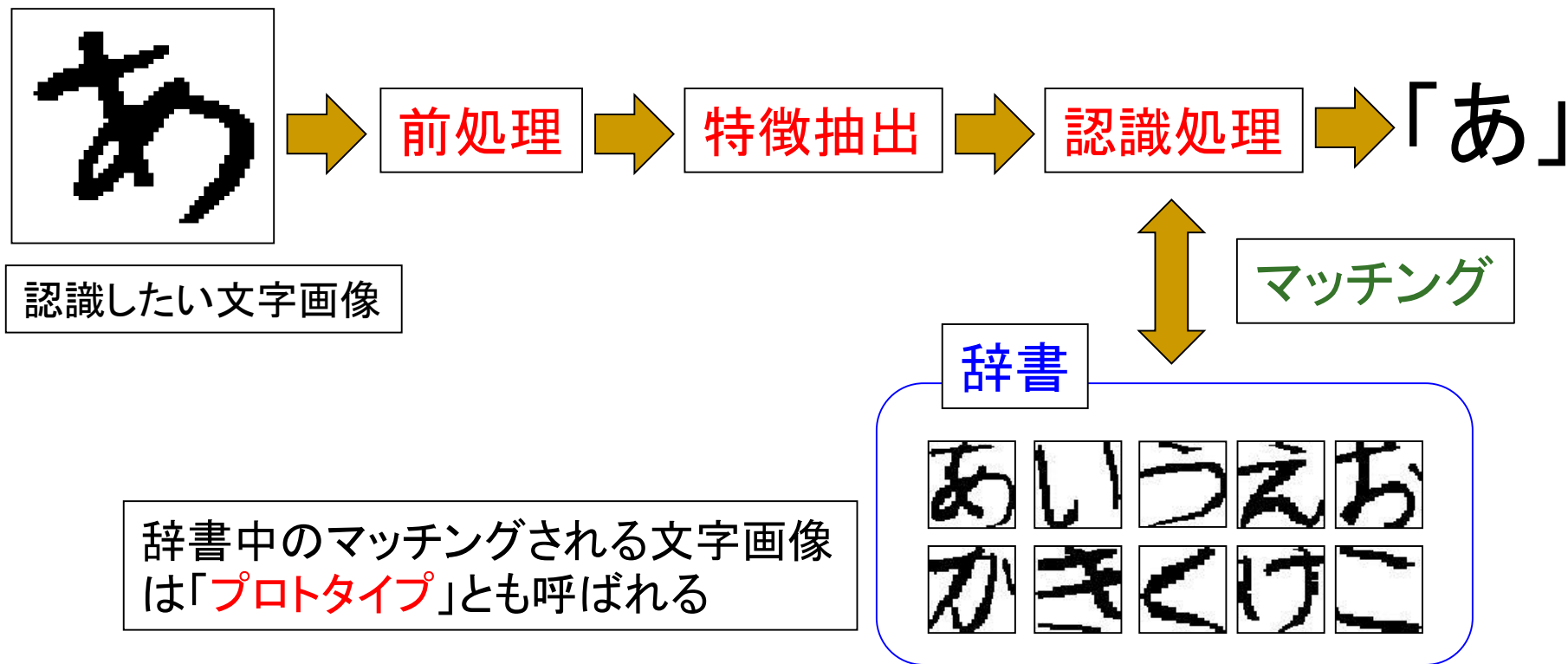
篠沢佳久

# 資料の内容

## ■ 統計的機械学習

- 決定境界による予測
  - 特徴の分布を用いた予測
  - ベイズ決定則
  - パラメータの推測(最尤推定)
- 
- 実習①(ベイズ決定則:表計算)
  - 実習②(ベイズ決定則:python)

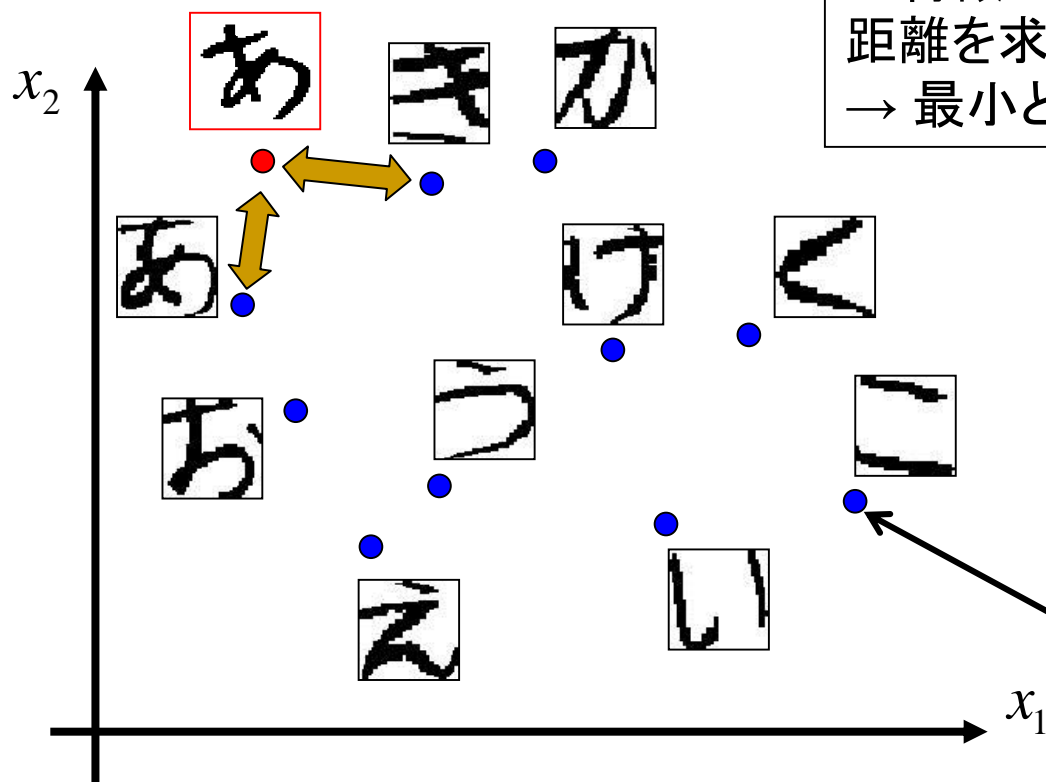
# 最近傍法



辞書内のプロトタイプとマッチングを行ない、最も類似度の高いものを結果とする方法を**最近傍法 (Nearest Neighbor)**と呼ぶ

# 最近傍法

認識したい文字画像  
(未知の文字画像)



特徴空間上において、認識したい文字画像と辞書の全ての文字画像(プロトタイプ)の特徴ベクトル(特徴空間上の座標)との距離を求める  
→ 最小となるものを認識結果\*

最近傍法  
(Nearest Neighbor)

プロトタイプ

\*類似度の場合は最大となるものを認識結果とします

# 最近傍法のアルゴリズム

$\text{max} = -\infty$

前処理(認識したい文字画像)

$\mathbf{t}$  = 特徴抽出(前処理後の文字画像)

for  $p$  in range( $\text{Max}_p$ ):

$\text{similarity} = \text{類似度}(\mathbf{t}, \mathbf{x}_p)$

    if  $\text{similarity} > \text{max}$ :

$\text{max} = \text{similarity}$

$\text{answer} = p$

$\text{Max}_p$  : プロトタイプの総数

$\mathbf{t}$  : 認識したい文字画像の特徴ベクトル

$\mathbf{x}_p$  : プロトタイプ  $p$  の特徴ベクトル

プロトタイプ  $\text{answer}$  が認識結果

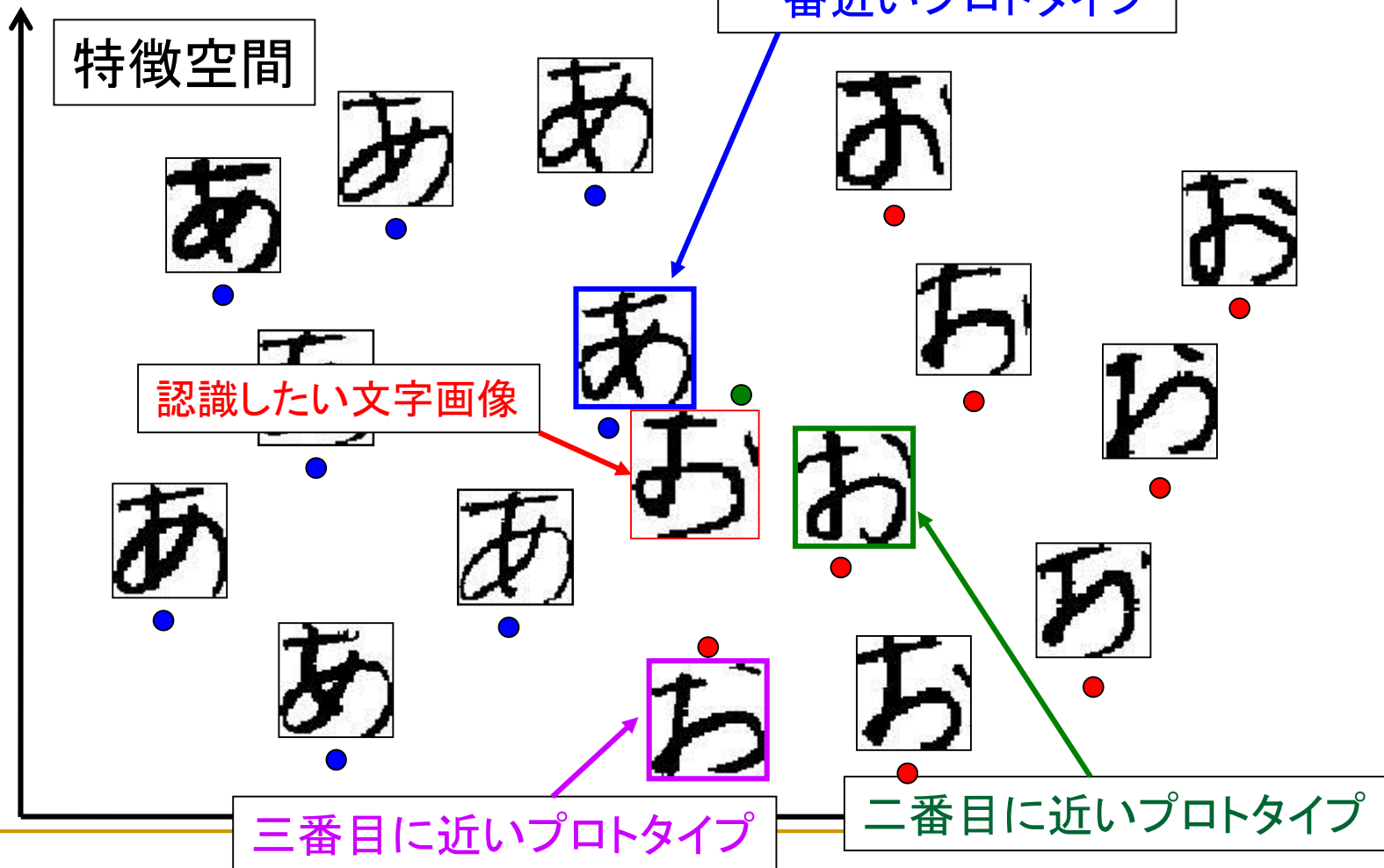
# k近傍法

- 一つの文字画像について複数個のプロトタイプを用意
- k番目までに近いプロトタイプを調べる
- k個の候補の多数決によって最終的な認識結果を決定
- $k=1$ の場合は、最近傍法と同等
- 問題点
  - プロトタイプ数の増加によって計算量が増加

# k近傍法

三番目まで近いプロトタイプを調べると「あ」が1個,「お」が2個  
→「お」と認識

k=3 の場合



# 統計的機械学習の基礎

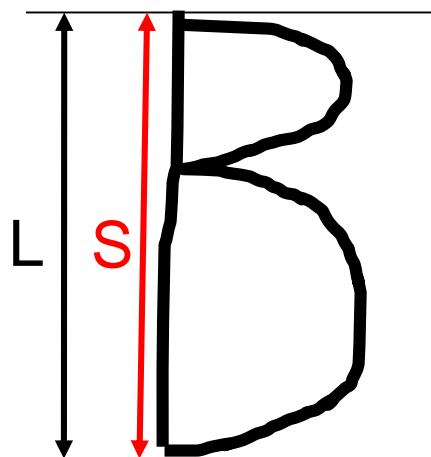
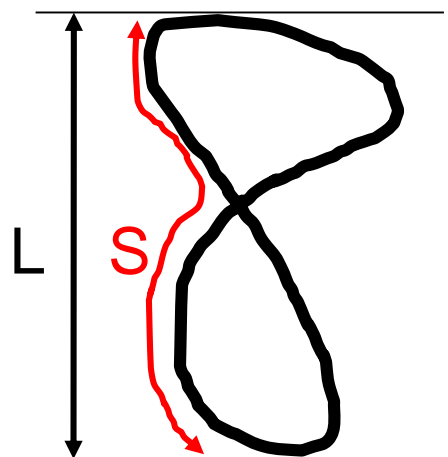
特徴の分布



# 特徴の一例

## ■ 「8」と「B」を区別する特徴

### □ 特徴量は一変数



直線を示す特徴

$$x = \frac{L}{S}$$

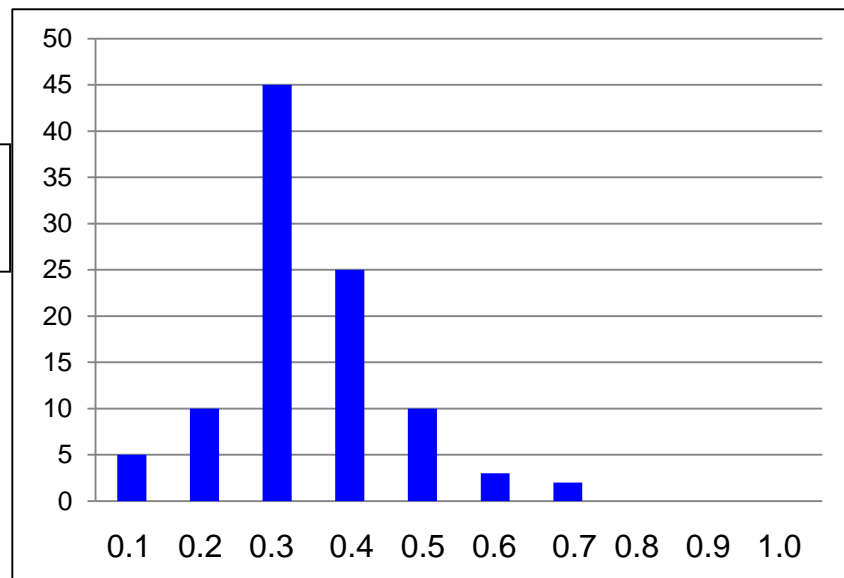
xが0に近い程 → 曲線的  
xが1に近い程 → 直線的

# 特徴の分布

「8」と「B」、それぞれ100個ずつの文字について直線を示す特徴の抽出を行なった場合

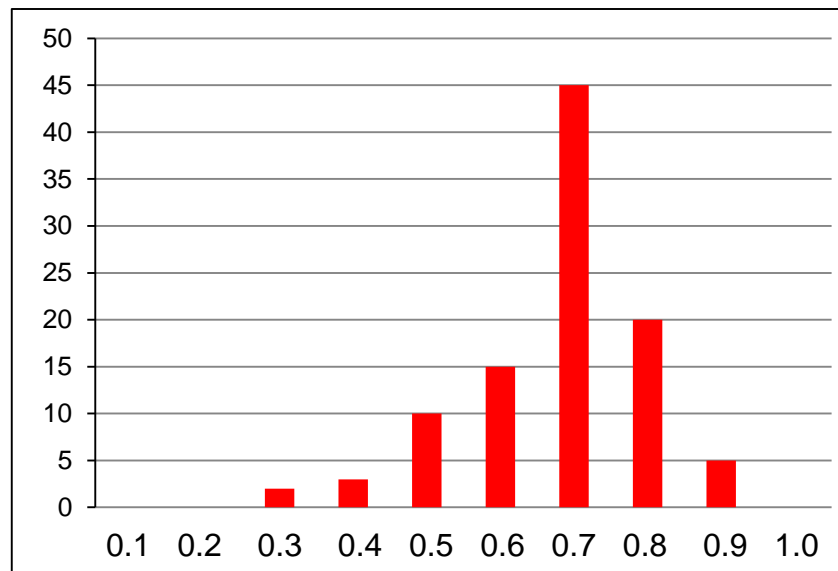
「8」の結果

度数



直線を示す特徴 x

「B」の結果



直線を示す特徴 x

# 特徴の出現確率(確率密度関数)①

## ■ $p(x|B)$

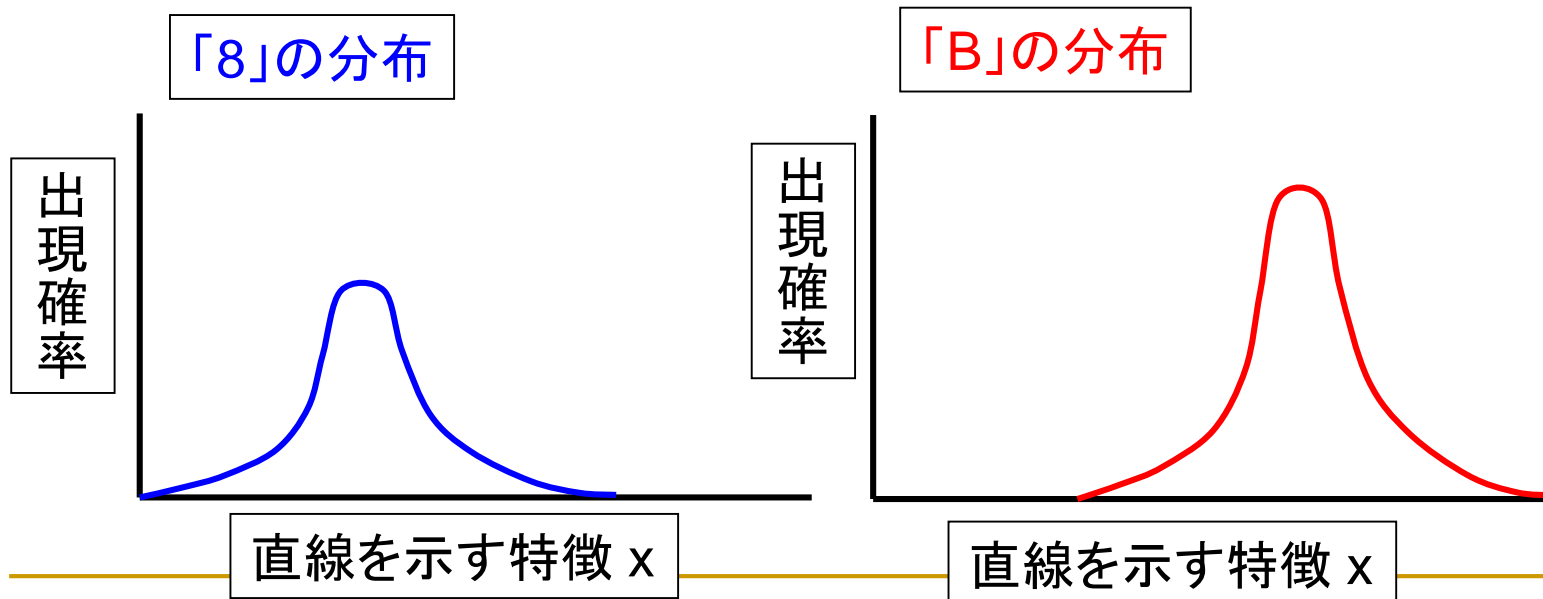
- 「B」という条件のもとで, 特徴  $x=0.3$  の出現する確率
- (例)  $p(x=0.3|B) = 45/100$

## ■ $p(x|8)$

- 「8」という条件のもとで, 特徴  $x=0.7$  の出現する確率
- (例)  $p(x=0.7|8) = 45/100$

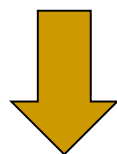
# 特徴の出現確率(確率密度関数)②

- 特徴の分布(確率密度関数)を知るためには？
  - 無限個の文字から特徴をとらなければならない
  - 現実的には不可能なため, 分布を仮定する
    - パラメータも推定する
  - 例えば正規分布(後ほど説明します)



# 特徴の分布からの認識

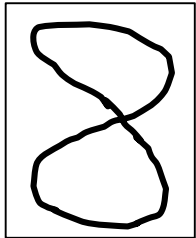
- 予測したい文字から特徴  $x$  を求め、「8」か「B」のどちらに属するのかを求めたい



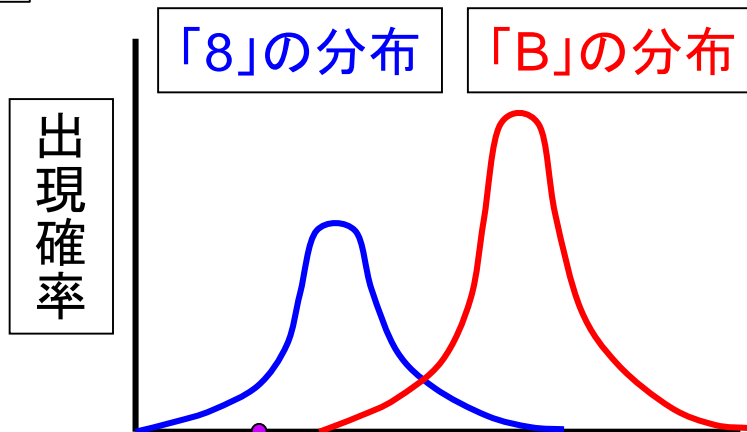
- 特徴の出現確率 ( $p(x|8)$ ,  $p(x|B)$ ) を利用する

統計的機械学習

# 統計的機械学習①



抽出した特徴 $x$ の値は 0.3



$x$  の値は0.3



$p(x|8) > p(x|B)$   
→ 「8」と予測

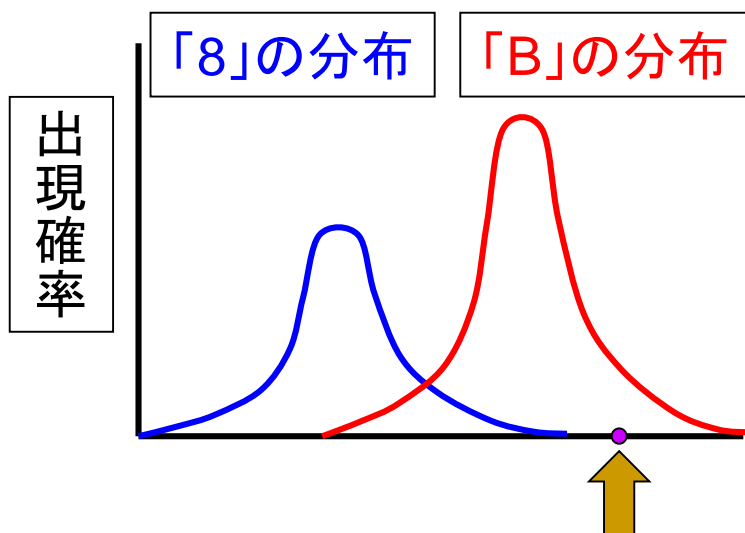
\*少々、正しくない説明をしていますが、今は気にしないで下さい

# 統計的機械学習①

B



抽出した特徴xの値は 0.8



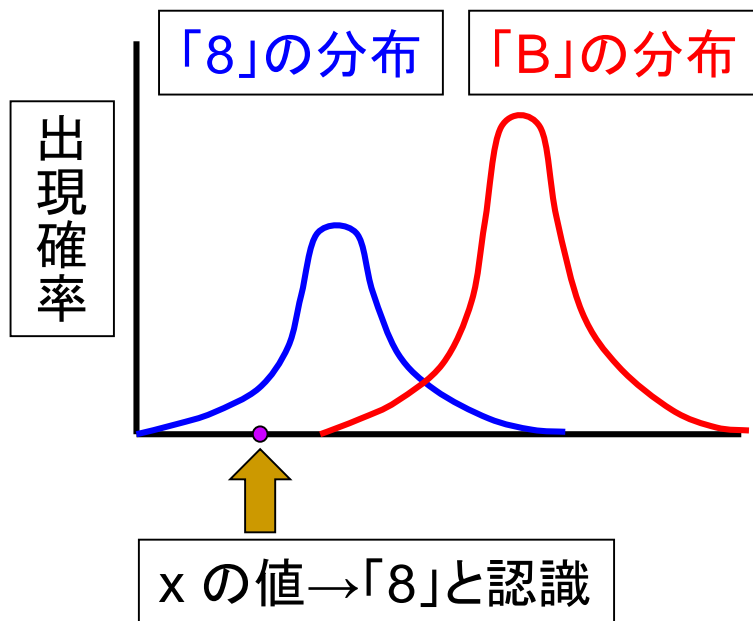
x の値は0.8



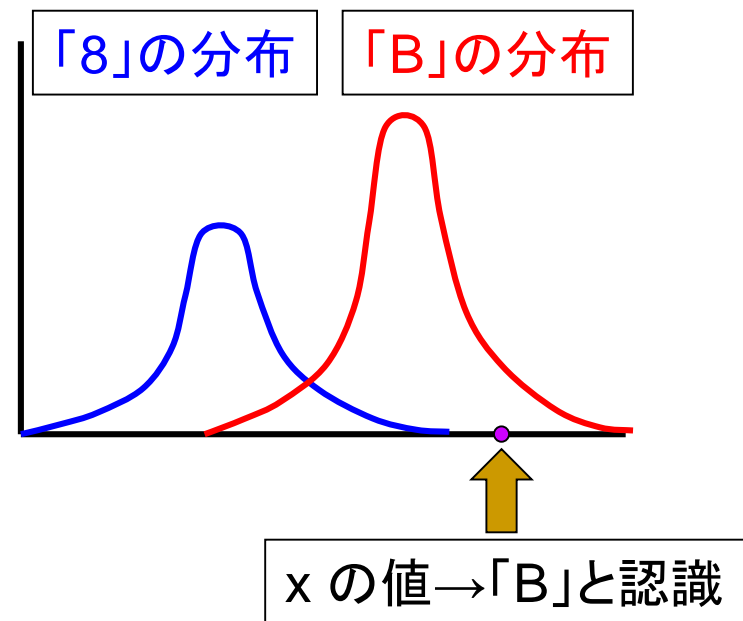
$p(x|8) < p(x|B)$   
→ 「B」と予測

# 統計的機械学習①

$p(x|8) > p(x|B)$  の場合  
→ 「8」と予測



$p(x|8) < p(x|B)$  の場合  
→ 「B」と予測

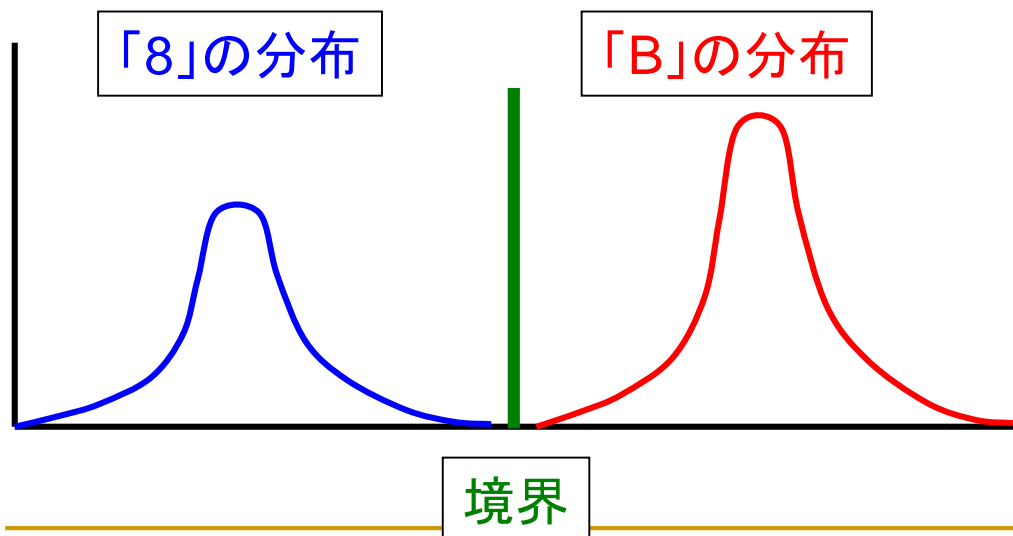




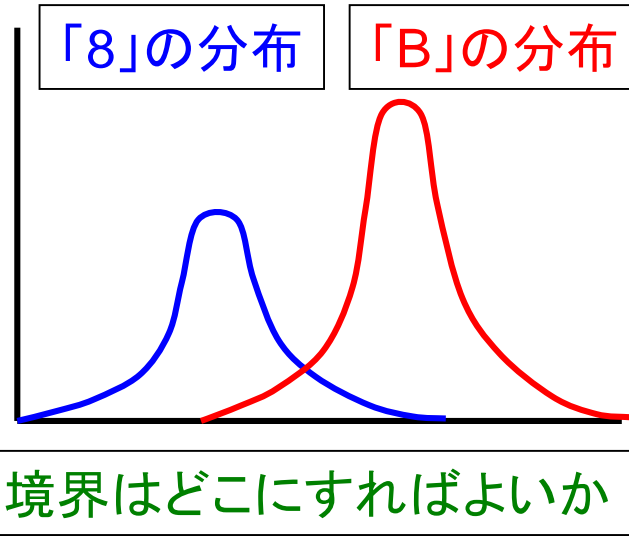
# 統計的機械学習②

- 二つの分布を分離する決定境界  $\tau$  を求める
  - 直線を示す特徴  $x > \tau \rightarrow$  「8」と予測
  - 直線を示す特徴  $x < \tau \rightarrow$  「B」と予測

分布が重ならない場合



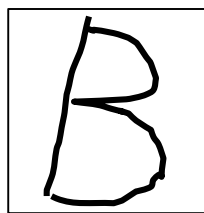
分布が重なっている場合



# 統計的機械学習③

## ■ $p(x|8)$

- 「8」という文字が与えられた場合,  $x$ の確率(条件付き確率\*)



$p(x|8)$ を使って比較するのは不自然\*\*

## ■ $p(8|x)$

- 「 $x$ という特徴が出現した場合, 「8」が出現する確率」を用いて比較するのが妥当
- ベイズ決定則

\* 尤度とも呼びます

\*\* 今回の資料のように  $p(8)=p(B)$  の場合は問題ない(後ほど説明します)

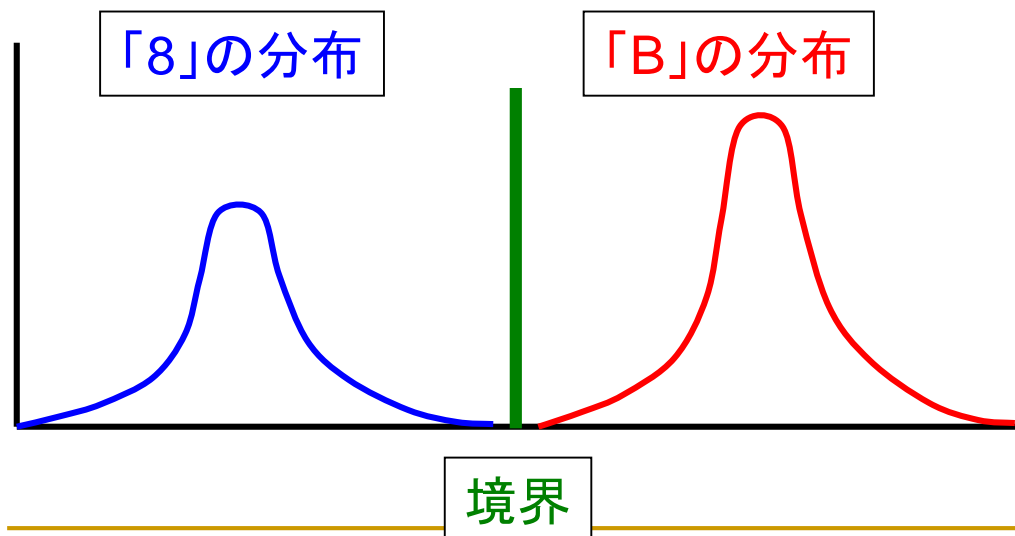
# 決定境界による予測 (統計的機械学習②)

ベイズ誤り

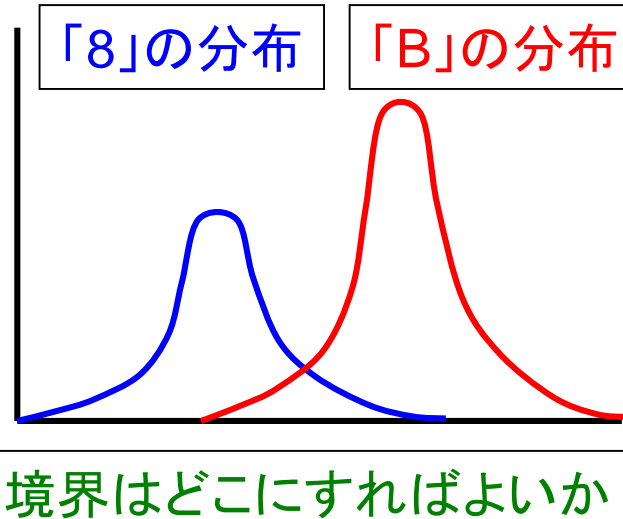
# 境界線の設定方法

- 二つの分布を分離する決定境界  $\tau$  を求める
  - 直線を示す特徴  $x > \tau \rightarrow$  「8」と予測
  - 直線を示す特徴  $x < \tau \rightarrow$  「B」と予測

分布が重ならない場合



分布が重なっている場合



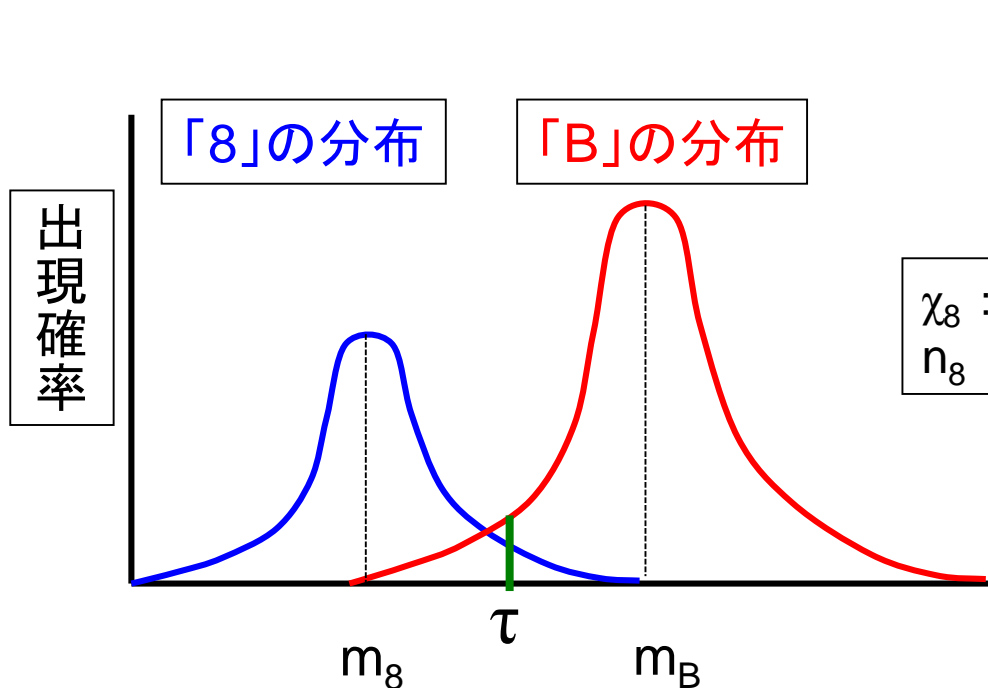
# 境界線の設定方法

- 平均値の平均値とする方法
- $p(\tau|8) = p(\tau|B)$  となる境界線  $\tau$  を求める方法
- 誤認識の割合を最小とする方法

# 境界の設定①

## ■ 設定方法①

- 「8」と「B」の特徴  $x$  の平均値  $m_8$ ,  $m_B$  とする
- 境界  $\tau = (m_8 + m_B) / 2$



$$m_8 = \frac{1}{n_8} \sum_{x \in \chi_8} x$$

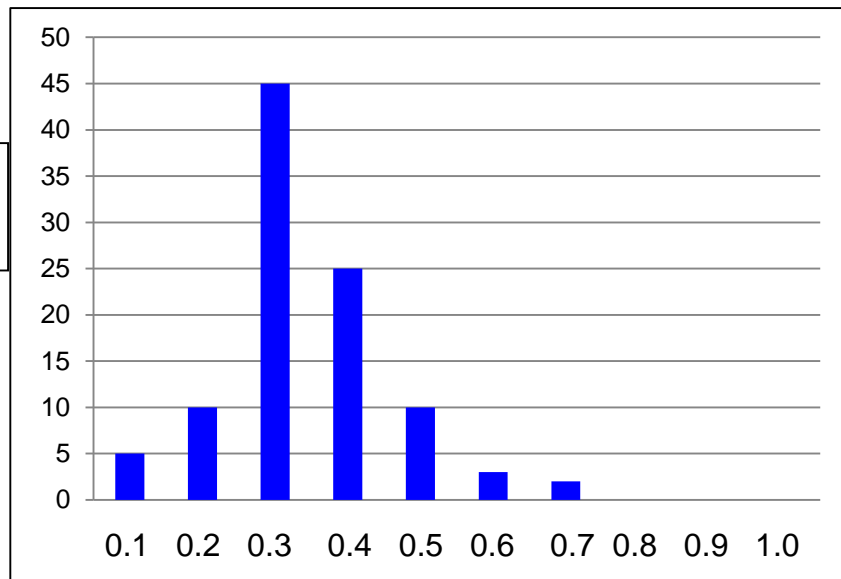
$\chi_8$  : 「8」に属する文字の集合  
 $n_8$  : 「8」に属する文字の総数

二つの平均値から境界を最も遠ざける

# 数値例(境界の設定方法①)

「8」の分布

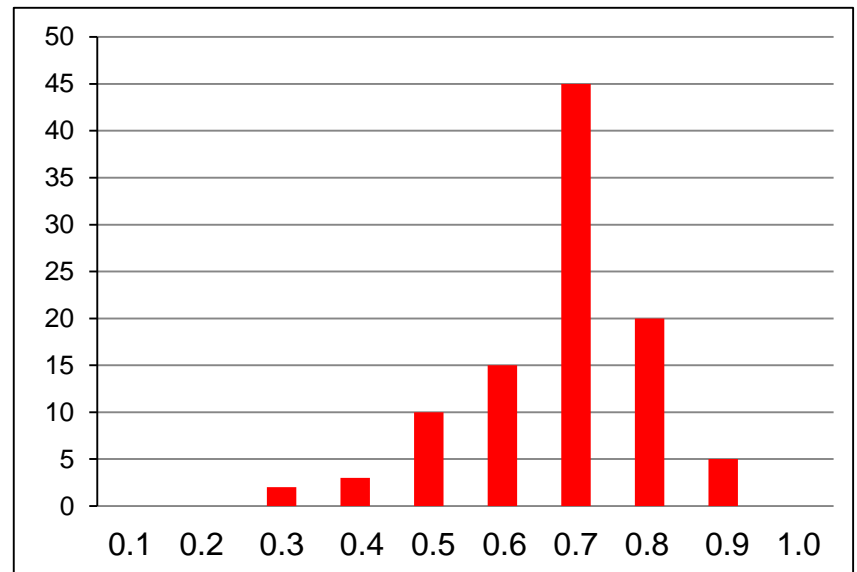
度数



直線を示す特徴 x

「8」の分布の平均値  
 $m_8=0.342$

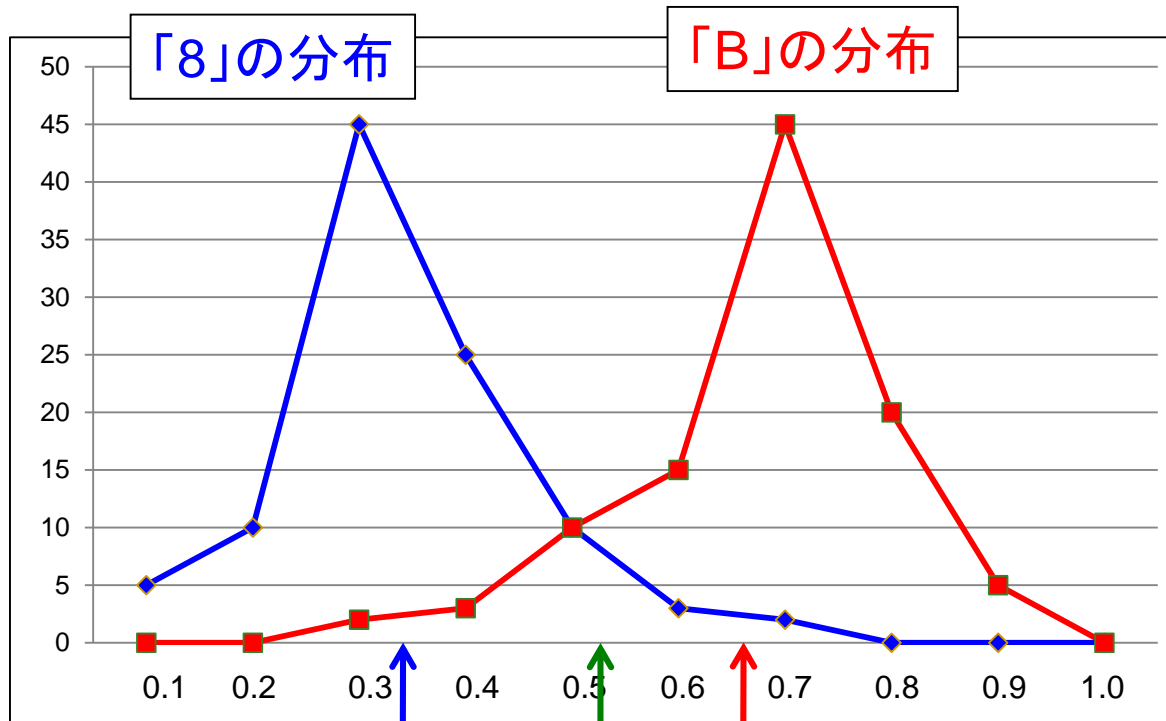
「B」の分布



直線を示す特徴 x

「B」の分布の平均値  
 $m_B=0.678$

# 数値例(境界の設定方法①)



8

$x = 0.4$  の場合  
 $\rightarrow x < \tau$   
 $\rightarrow$  「8」と予測

B

$x = 0.8$  の場合  
 $\rightarrow x > \tau$   
 $\rightarrow$  「B」と予測

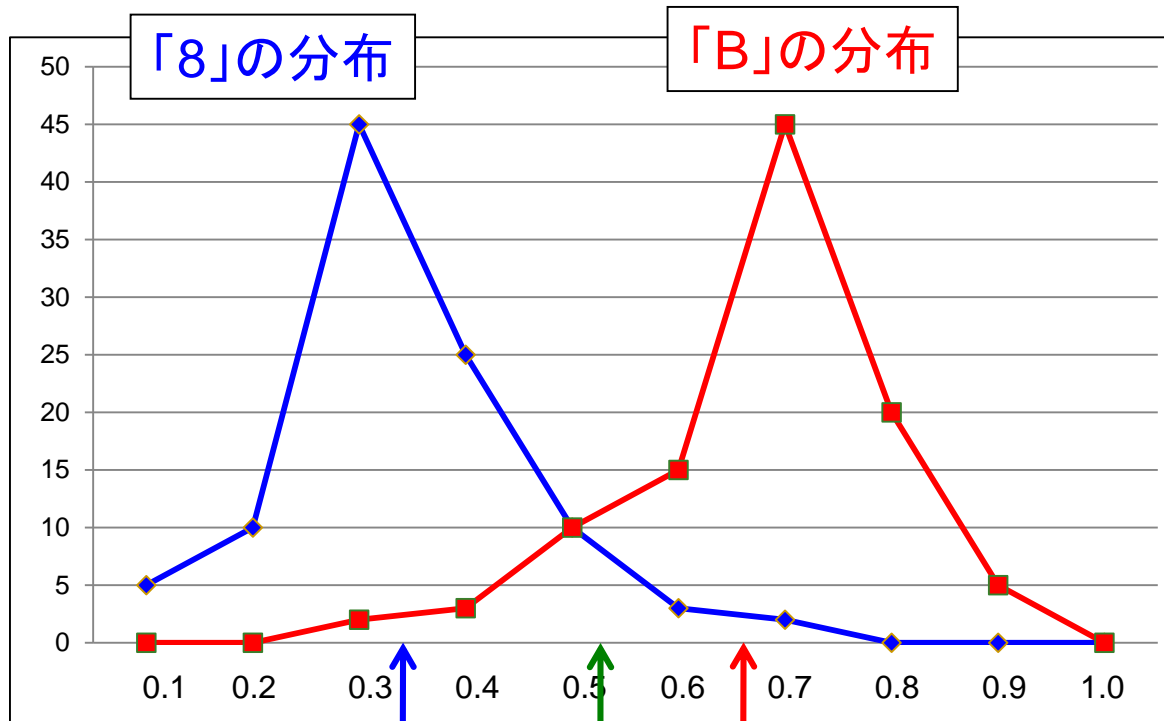
「8」と予測

「B」と予測

境界  $\tau = (0.342 + 0.678) / 2 = 0.51$



# 数値例(境界の設定方法①)



8

$x = 0.5$  の場合  
 $\rightarrow x < \tau$   
 $\rightarrow$  「8」と予測

B

$x = 0.5$  の場合  
 $\rightarrow x < \tau$   
 $\rightarrow$  「8」と予測  
(誤認識)

「8」と予測

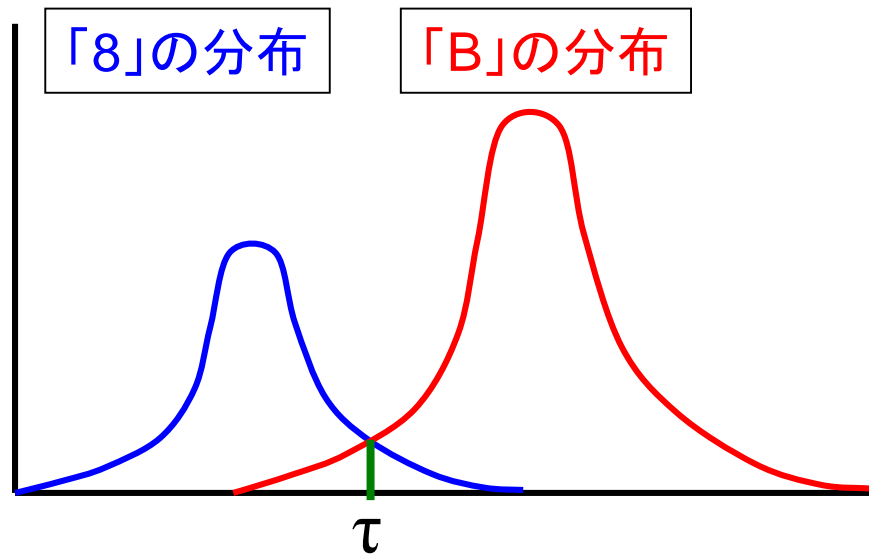
「B」と予測

境界  $\tau = (0.342 + 0.678) / 2 = 0.51$

# 境界の設定②

## ■ 設定方法②

- $p(\tau|8) = p(\tau|B)$  となる境界線  $\tau$  を求める

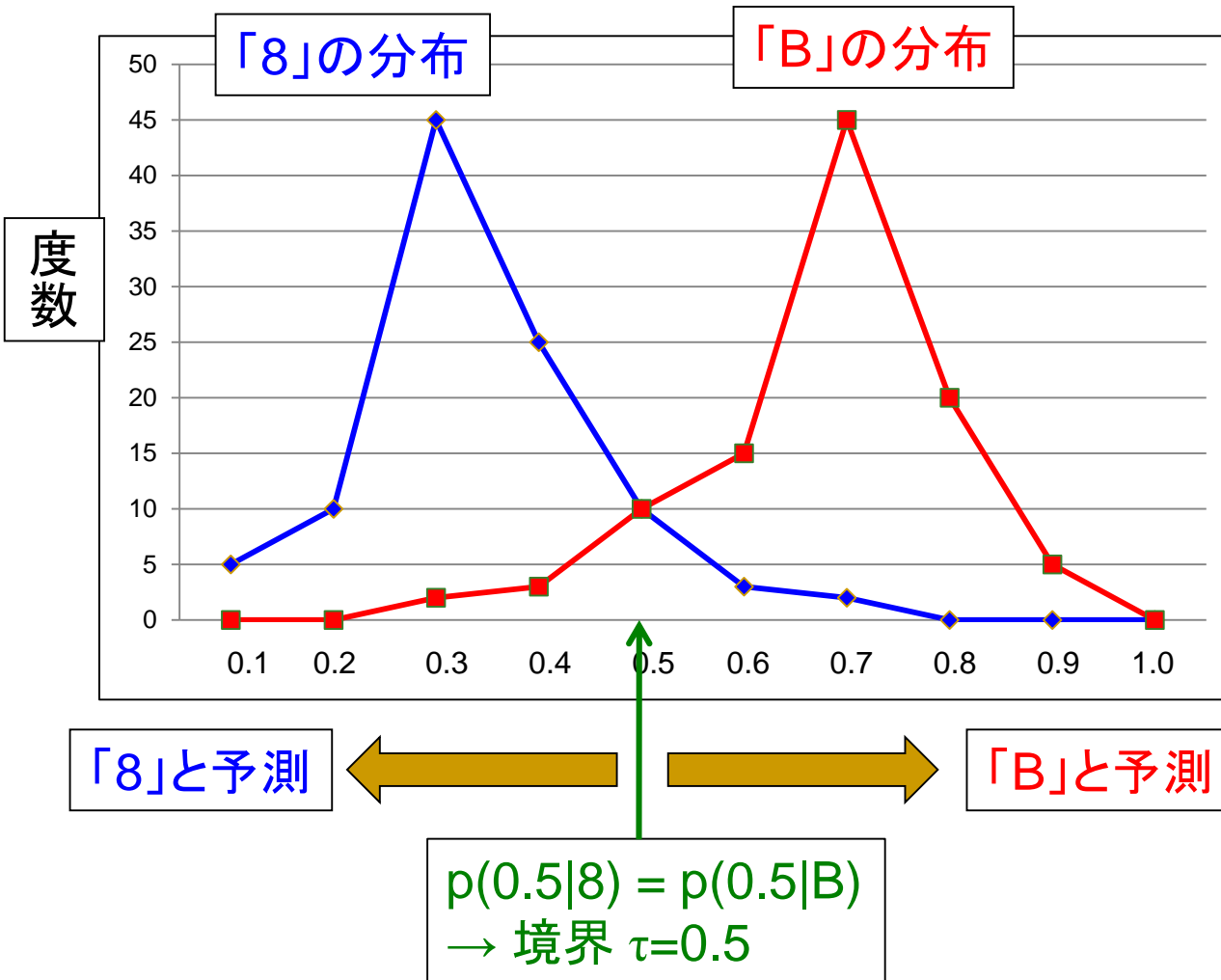


$\tau > x$  の場合  
 $\rightarrow p(\tau|8) > p(\tau|B)$

$\tau < x$  の場合  
 $\rightarrow p(\tau|8) < p(\tau|B)$

$$p(\tau|8) = p(\tau|B)$$

# 数値例(境界の設定方法②)



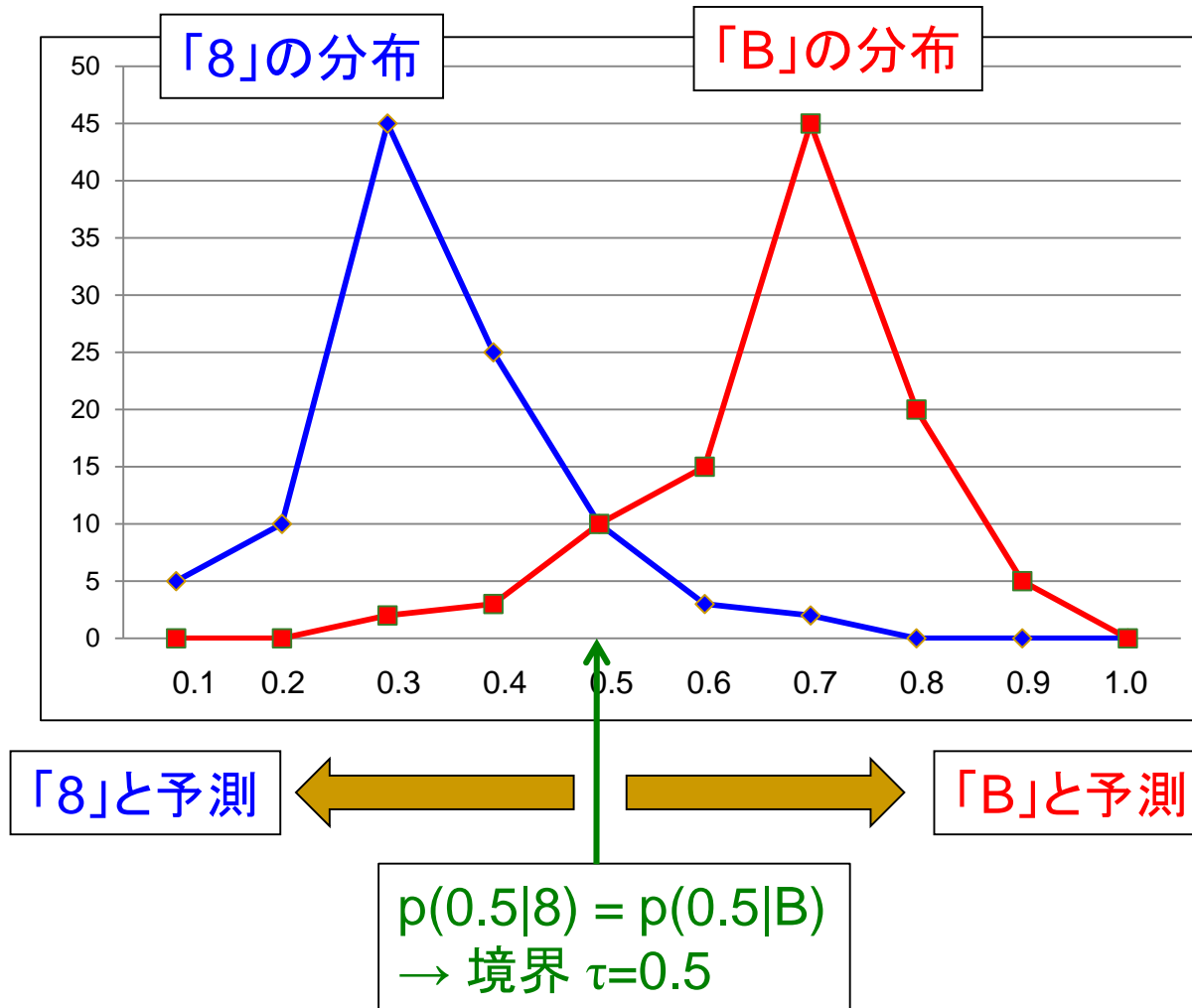
8

$x = 0.4$  の場合  
→  $x < \tau$   
→ 「8」と予測

B

$x = 0.8$  の場合  
→  $x > \tau$   
→ 「B」と予測

# 数値例(境界の設定方法②)



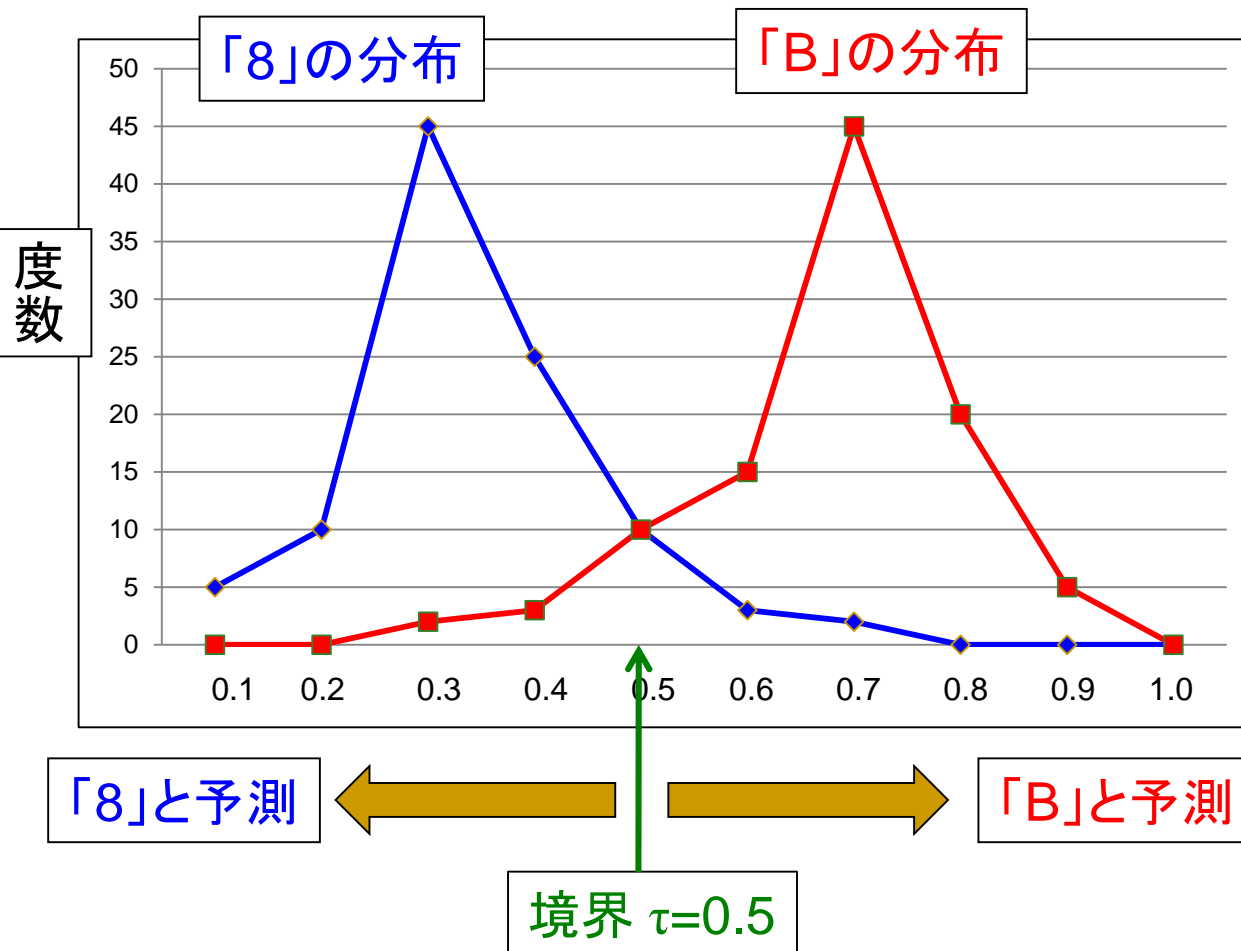
8

$x = 0.5$  の場合  
→  $x = \tau$   
→ 予測できない

B

$x = 0.5$  の場合  
→  $x = \tau$   
→ 予測できない

# 誤認識の可能性



境界  $\tau=0.5$  とした場合

「8」において, 0.5以上の  
特徴が出現する確率  
は0.15  
→ 0.15 の割合で「8」を  
「B」と誤認識する可能  
性がある

「B」において, 0.5以下  
の特徴が出現する確率  
は0.15  
→ 0.15 の割合で「B」を  
「8」と誤認識する可能  
性がある

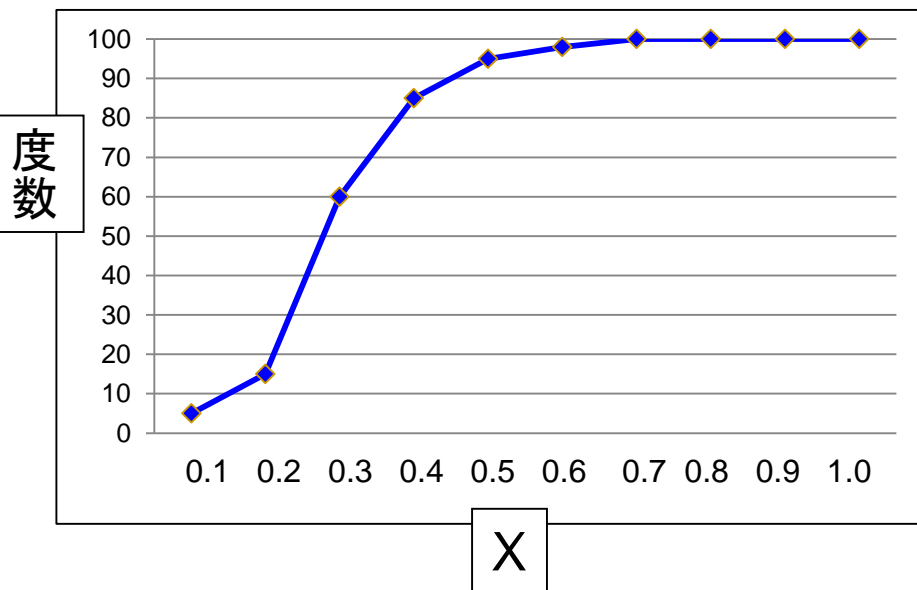
# 境界の設定③

## ■ 周辺累積分布

- 特徴  $x$  の値が  $X$  以下である確率

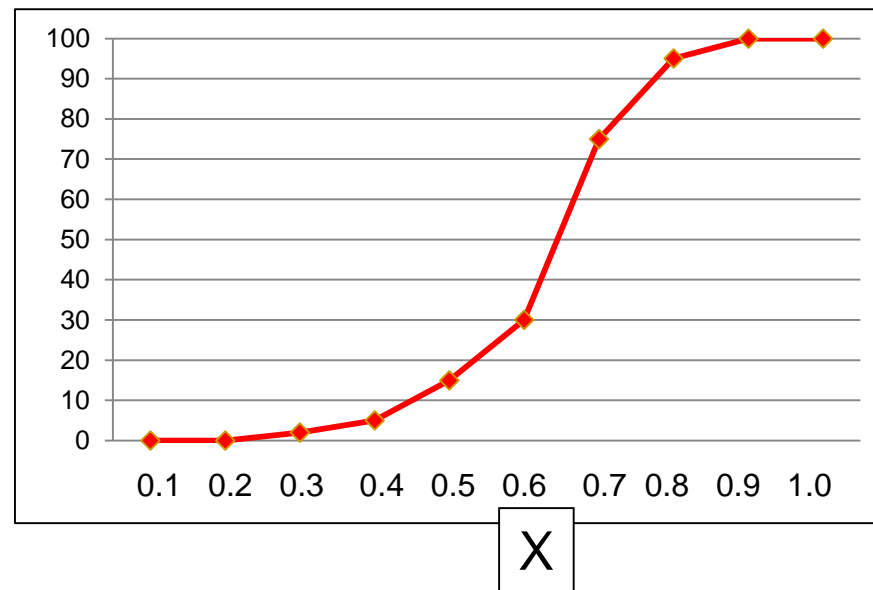
「8」の分布

$$F(X | 8) = \int_{-\infty}^X p(x | 8) dx$$



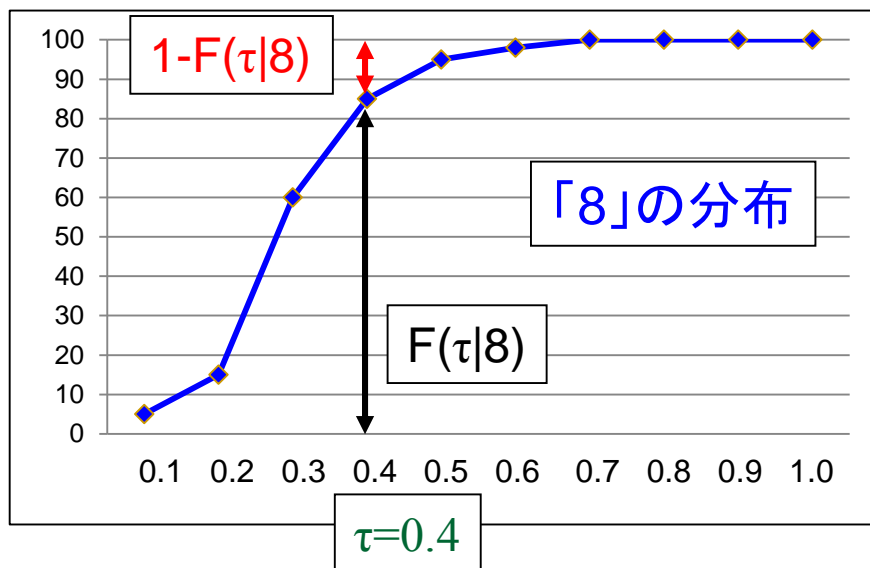
「B」の分布

$$F(X | B) = \int_{-\infty}^X p(x | B) dx$$



# 境界の設定③

## ■ 「8」の周辺累積分布



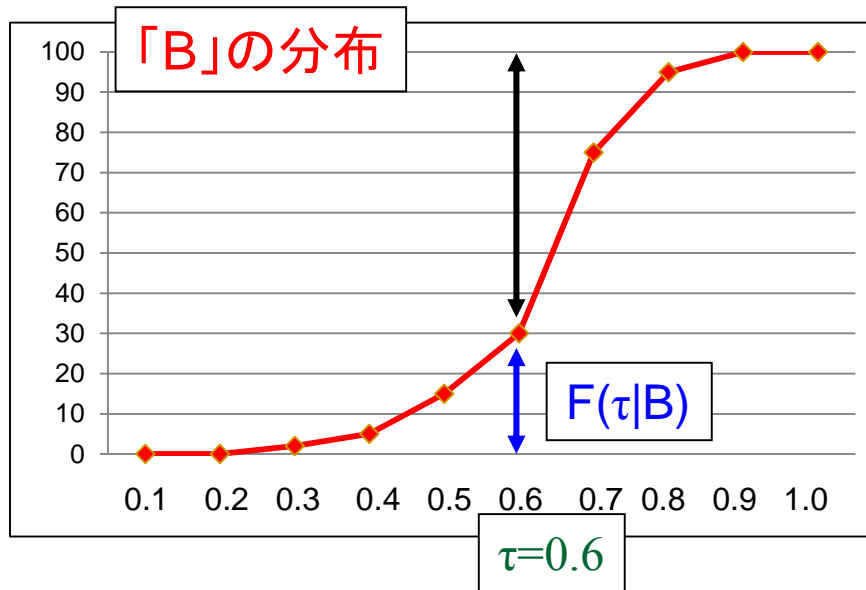
境界に  $\tau$  を設定した場合  
 $1-F(\tau|8)$   
「8」を「B」と誤認識する割合

境界  $\tau = 0.4$  と設定した場合

- 特徴  $x$  が 0.4 以下の「8」は  $85/100=0.85$  の割合で出現
- 残り 0.15 の割合で、0.4 よりも大きい特徴の「8」が出現する可能性がある
- 0.15 の割合で、「8」を「B」と間違える可能性がある

# 境界の設定③

## ■ 「B」の周辺累積分布



境界に  $\tau$  を設定した場合  
 $F(\tau|B)$   
「B」を「8」と誤認識する割合

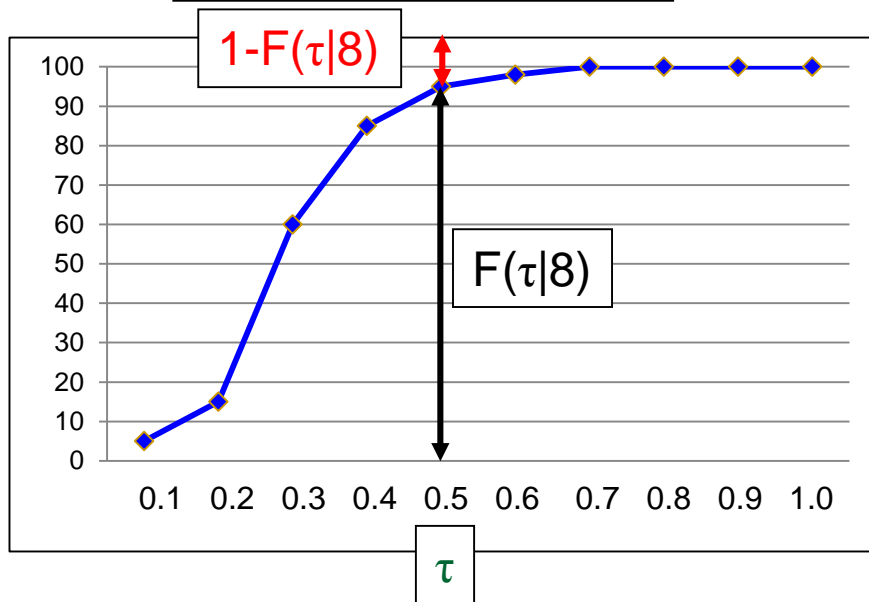
境界  $\tau = 0.6$  と設定した場合

- 特徴  $x$  が 0.6 以上の「B」は  $70/100=0.7$  の割合で出現
- 残り 0.3 の割合で、0.6 よりも小さい特徴の「B」が出現する可能性がある
- 0.3 の割合で、「B」を「8」と間違える可能性がある

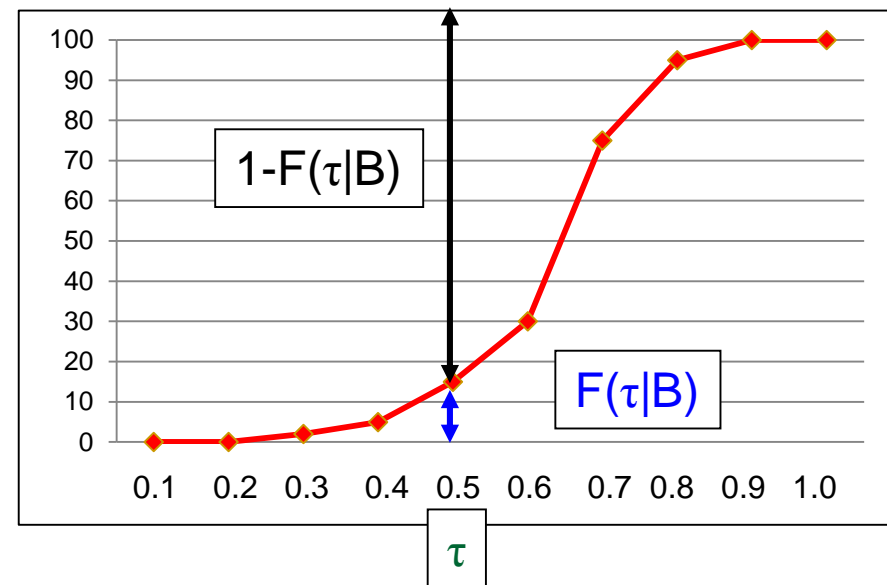


# 境界の設定③

「8」の周辺累積分布



「B」の周辺累積分布



- 誤認識の割合を最小
- $1 - F(\tau|8) + F(\tau|B) \rightarrow$  最小となる境界  $\tau$  を求める

# 境界の設定③

$$1-F(\tau|8)$$

「8」を「B」と誤認識する割合

## ■ 境界の設定方法③

□ 誤認識の割合を最小

□  $1-F(\tau|8)+F(\tau|B) \rightarrow$  最小となる境界  $\tau$  を求める



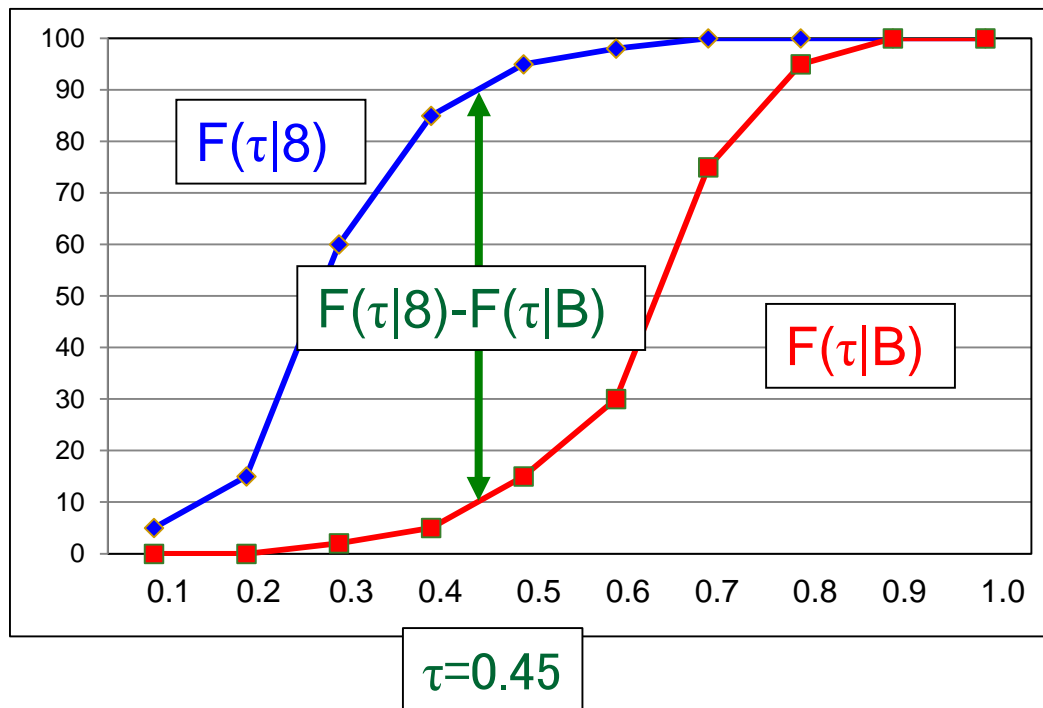
□  $F(\tau|8)-F(\tau|B) \rightarrow$  最大となる境界  $\tau$  を求める

$$F(\tau|B)$$

「B」を「8」と誤認識する割合

# 数値例(境界の設定方法③)

$F(0.45|8)-F(0.45|B)$  が最大  
→ 境界  $\tau=0.45$

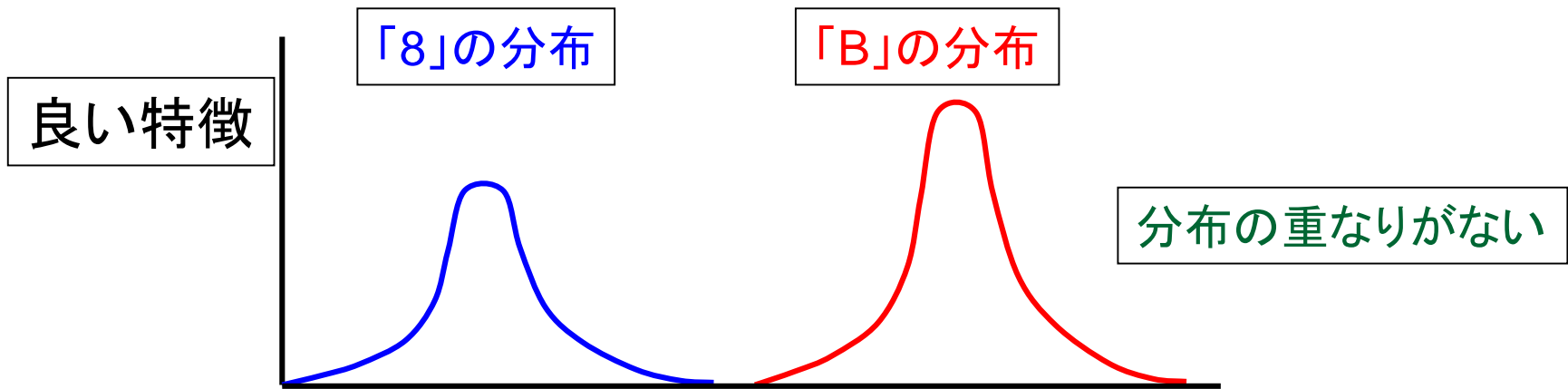
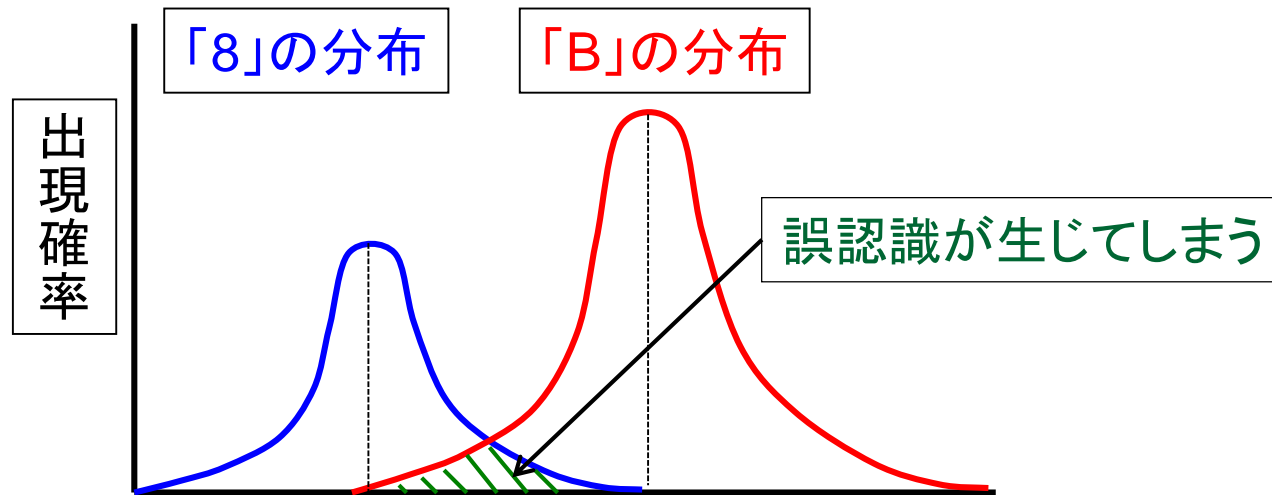


# 分布の重なりによる誤認識①

## ■ 必然的な誤り

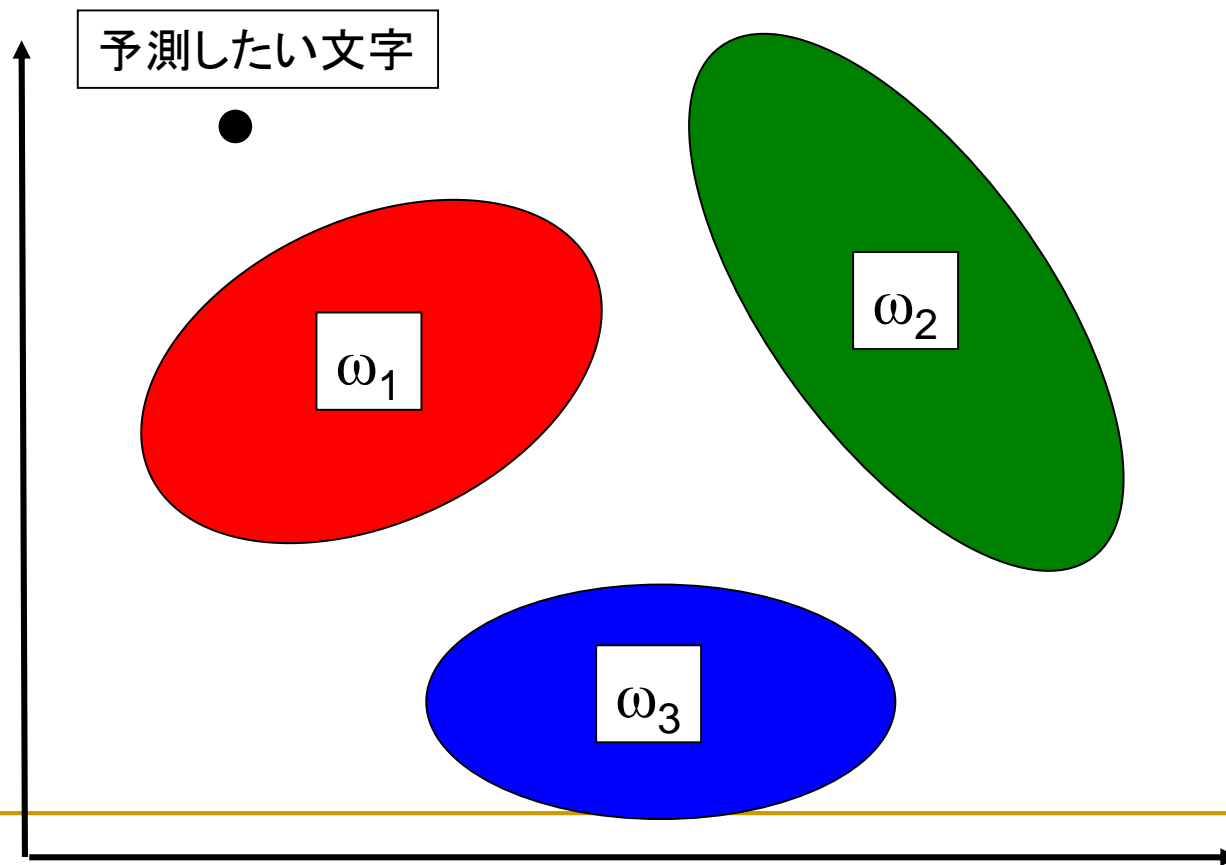
- 境界の設定によって、誤認識が生じる場合がある
- 分布の重なりによって、必然的な誤り(ベイズ誤り)が生じてしまう
- 必然的な誤りが少ない特徴が良い特徴とされる

# 分布の重なりによる誤認識②



# クラスが複数個の場合①

- クラス数が3個の場合
  - 特徴ベクトルは2次元

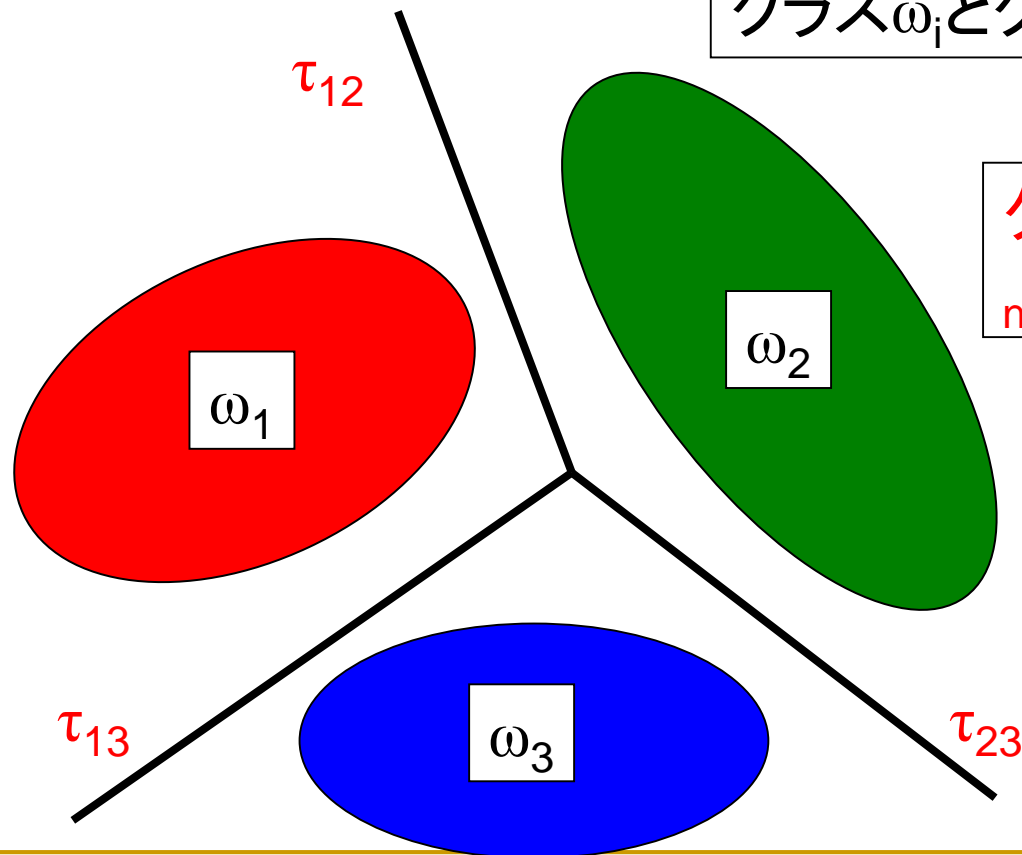


## クラスが複数個の場合②

- 「1対1」(one-versus-one)による方法
- 「1対その他」(one-versus-rest)による方法

# 「1対1による方法」①

- クラス数が3個の場合
  - 特徴ベクトルは2次元

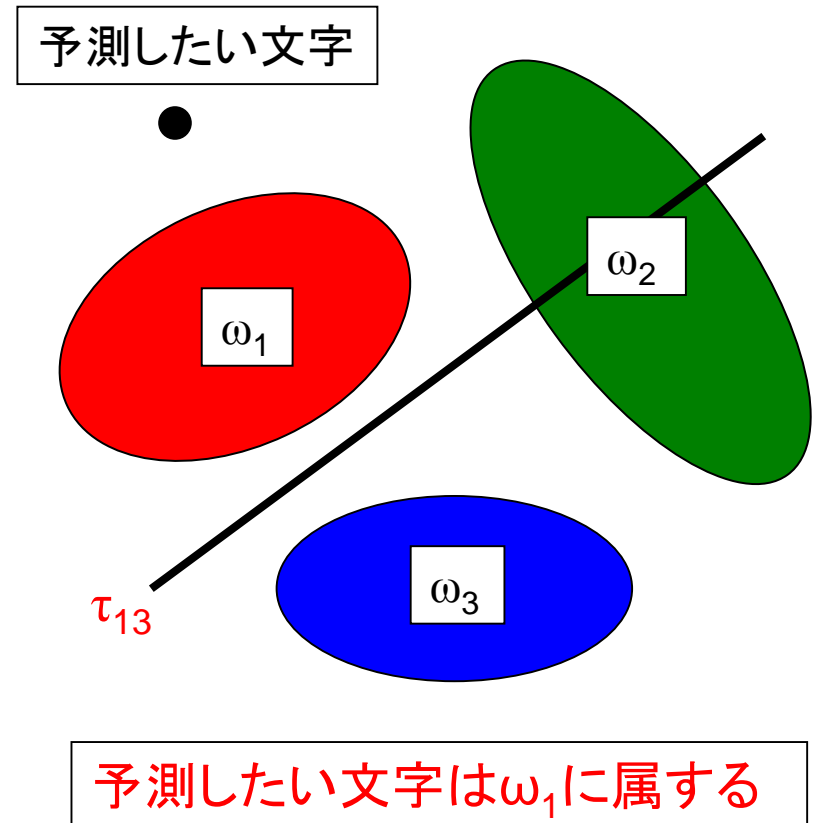
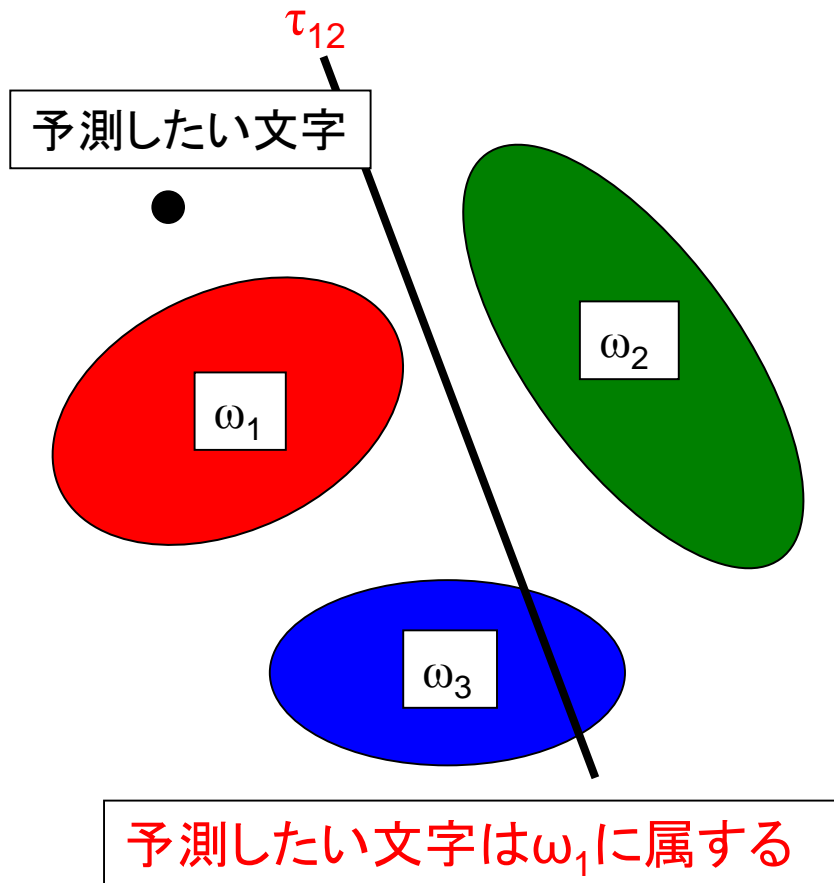


$\tau_{ij}$   
クラス $\omega_i$ とクラス $\omega_j$ を分ける境界

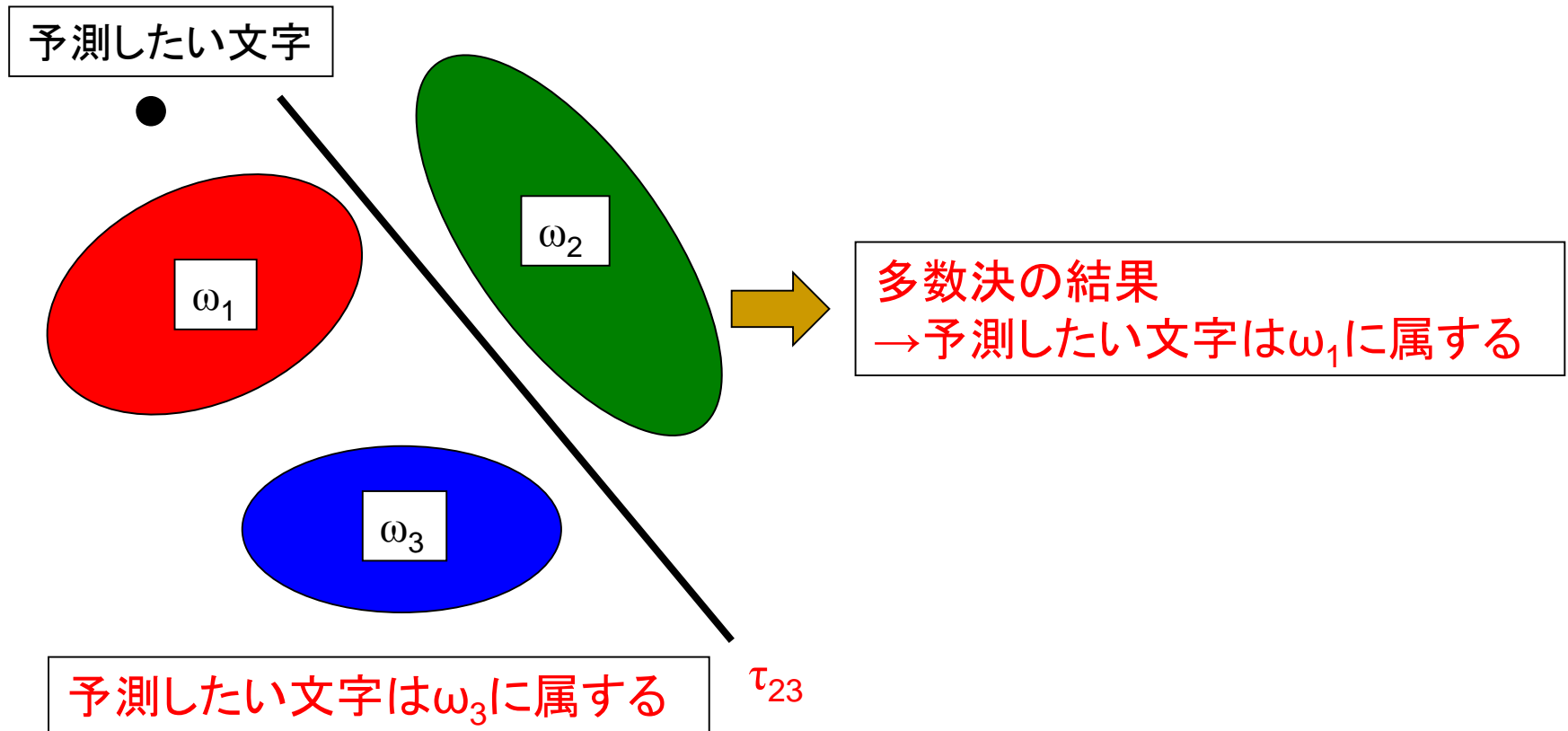
クラス数が $n$ 個の場合  
 $nC_2$ の境界



# 「1対1による方法」②



# 「1対1による方法」③



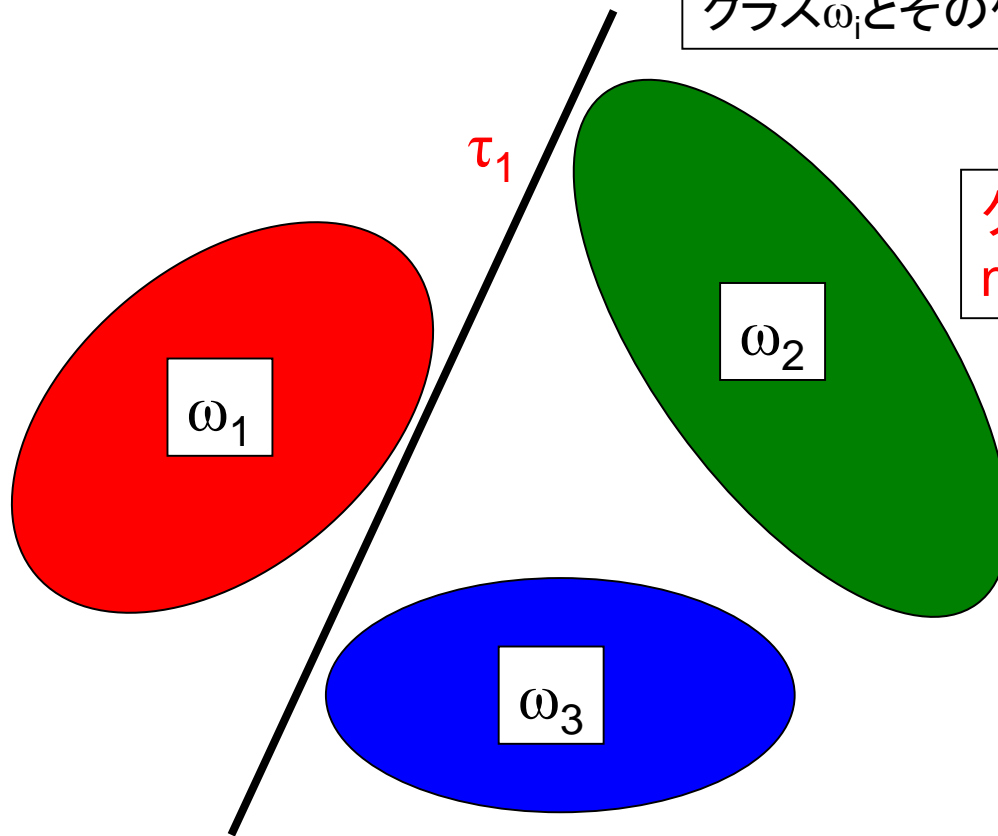
# 「1対1による方法」④

## ■ クラス数が $n$ 個の場合

- 二つの文字を分離する境界線を ${}_nC_2$ 個求める
- 予測したい文字がどちらに属するか,  ${}_nC_2$ 個の境界線において調べる
- クラス数が多くなった場合, 解に矛盾が生じる
- 多数決によって最終的に解を決める場合もある

# 「1対その他による方法」①

- クラス数が3個の場合
  - 特徴ベクトルは2次元

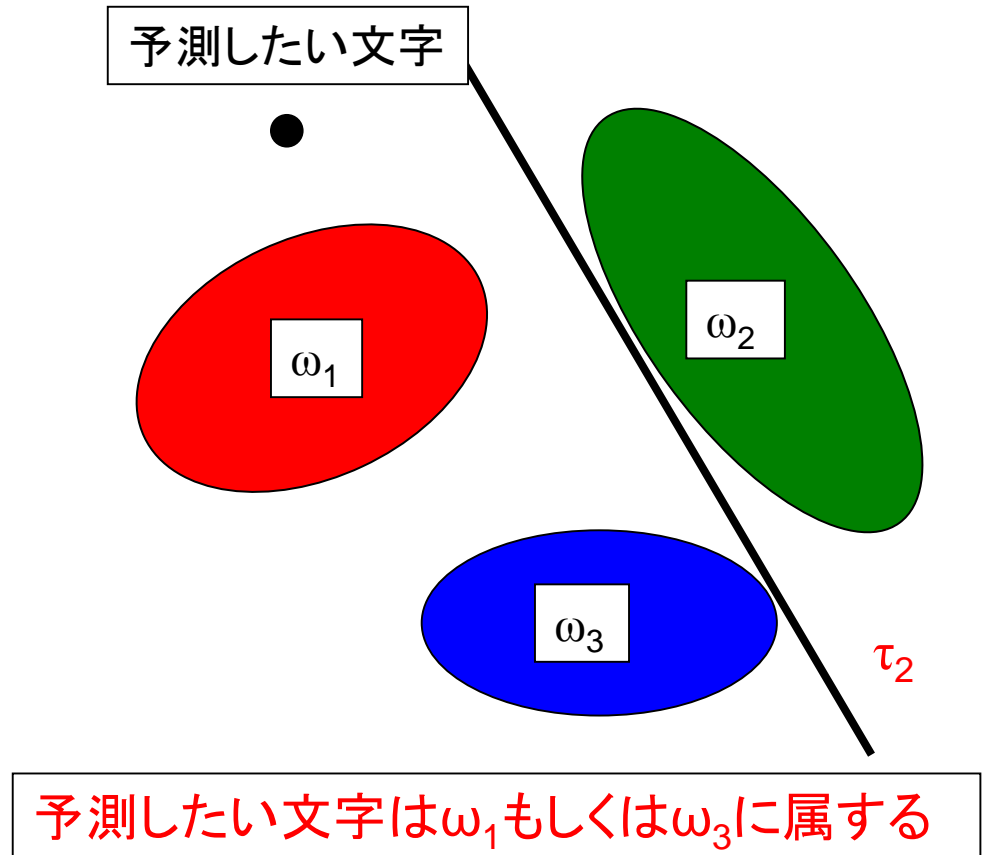
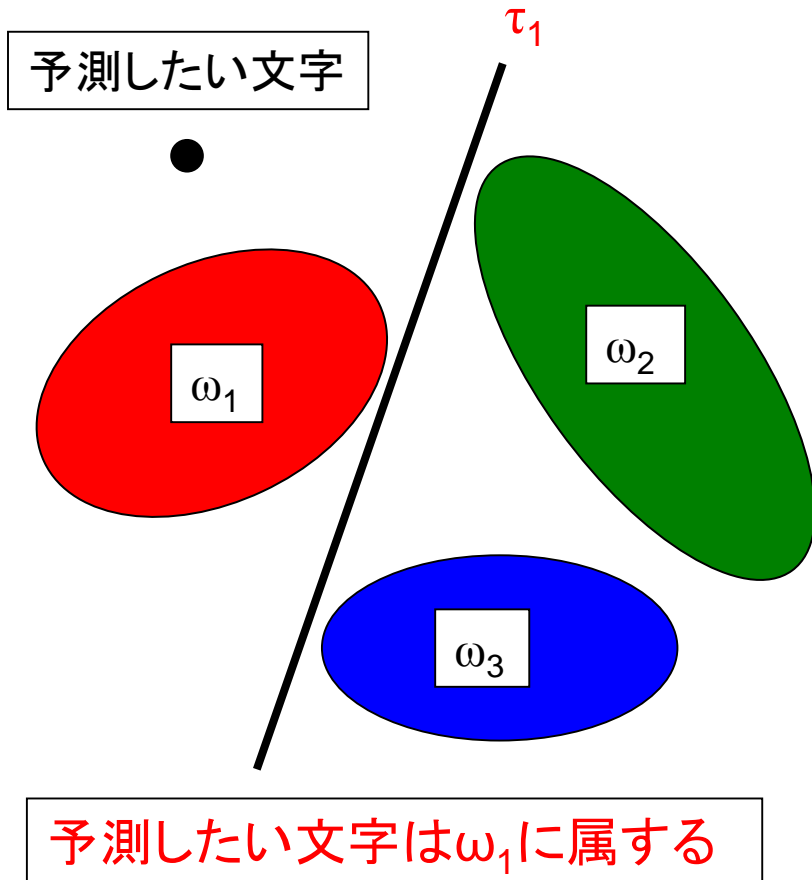


$\tau_i$

クラス $\omega_i$ とその他のクラスを分ける境界

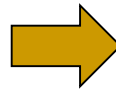
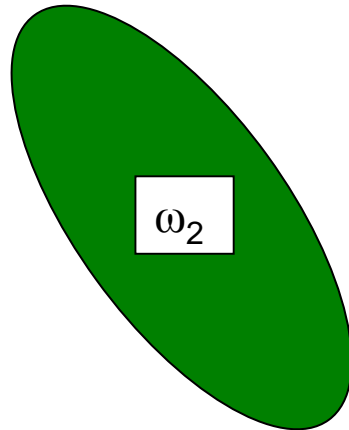
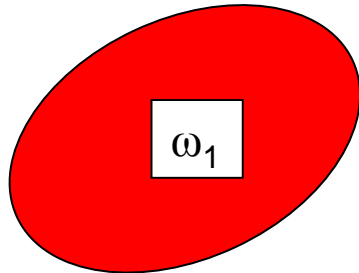
クラス数が $n$ 個の場合  
 $n$ 個の境界

# 「1対その他による方法」②

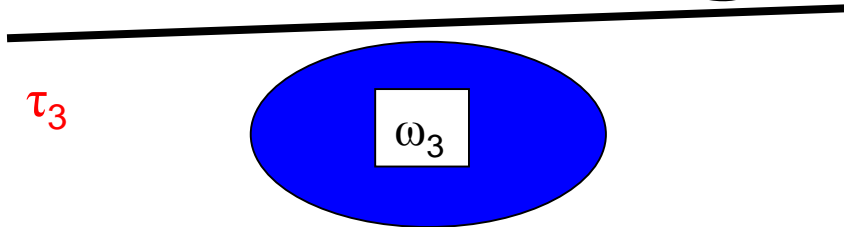


# 「1対その他による方法」③

予測したい文字



多数決の結果  
→ 予測したい文字は  $\omega_1$  に属する



予測したい文字は  $\omega_1$  にもしくは  $\omega_2$  に属する

# 「1対その他による方法」④

## ■ クラス数が $n$ 個の場合

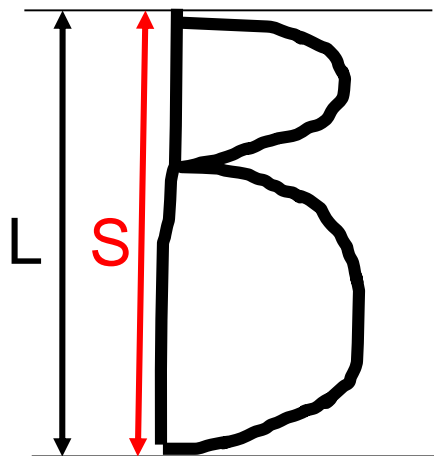
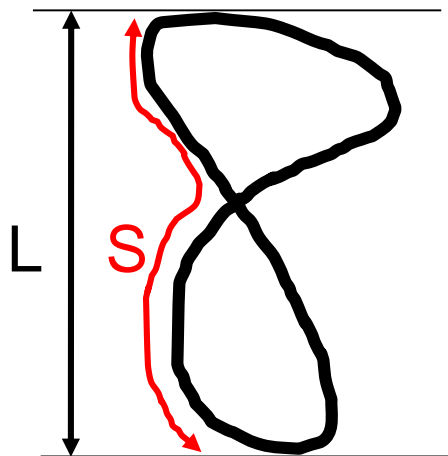
- 任意のクラスとその他のクラスを分離する境界線を $n$ 個求める
- 予測したい文字がどちらに属するか,  $n$ 個の境界線において調べる
- クラス数が多くなった場合, 解に矛盾が生じる
- 多数決によって最終的に解を決める場合もある

# 出現確率を用いた予測方法 (統計的機械学習①)

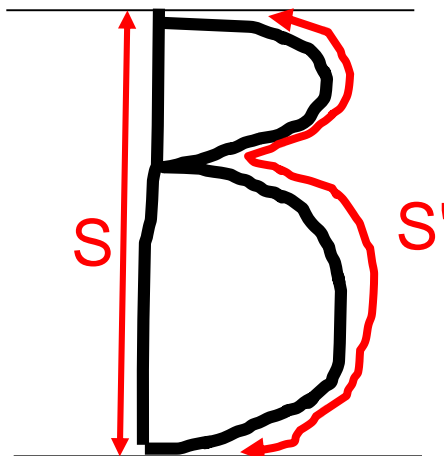
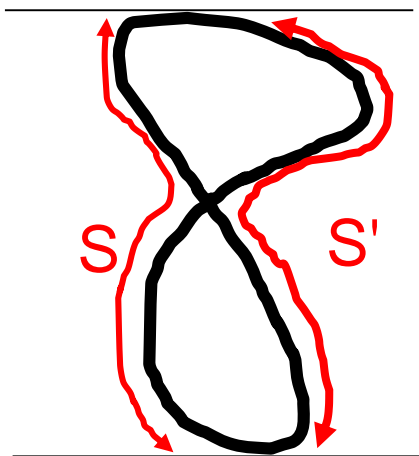
正規分布  
マハラノビス距離



# 新しい特徴抽出



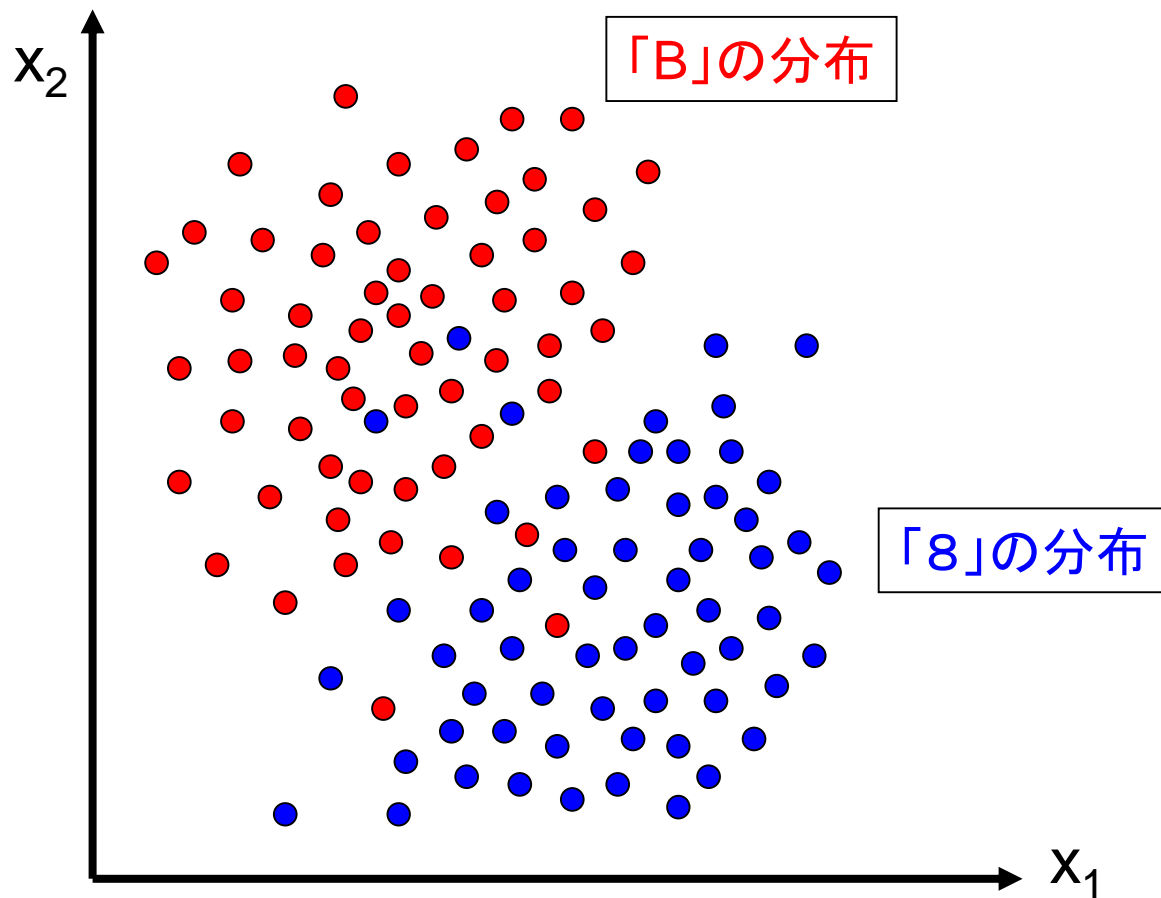
$$x_1 = \frac{L}{S}$$



$$x_2 = \frac{S}{S'}$$

# 特徴空間

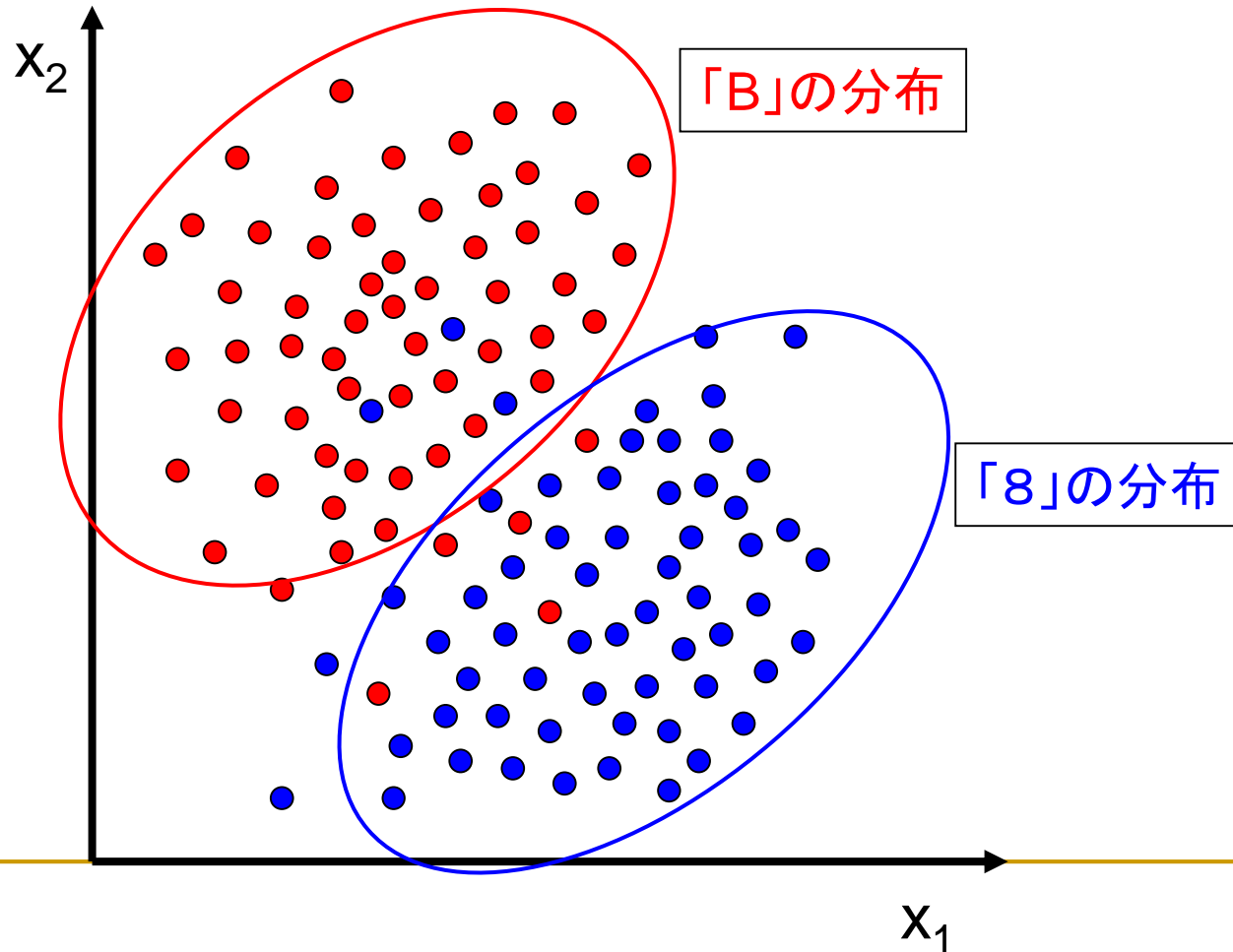
- 「8」と「B」の特徴ベクトル( $x_1, x_2$ )を平面上に表現



# クラスター

## ■ クラスター

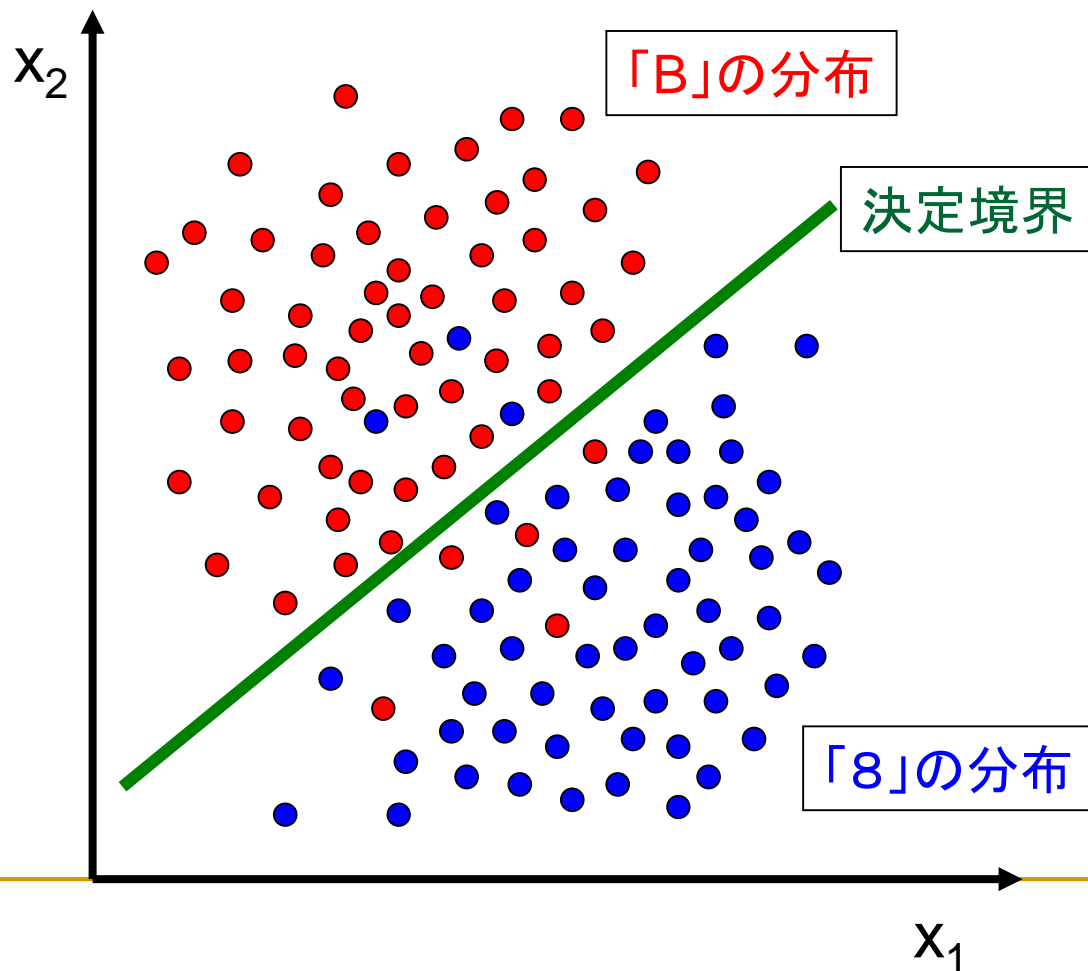
- 特徴空間上において、クラスごとにまとまって観測される塊



# 決定境界

## ■ 決定境界

### □ クラス間を分離する境界



# 特徴の分布の仮定①

## ■ 特徴の確率密度関数

- 特徴  $x_i (i=1,2)$  の出現確率は一般的には分からない
- $x_i (i=1,2)$  の出現確率は**独立**と仮定する



$$p(\mathbf{x} | 8) = p(x_1, x_2 | 8) = p(x_1 | 8) p(x_2 | 8)$$

$$p(\mathbf{x} | B) = p(x_1, x_2 | B) = p(x_1 | B) p(x_2 | B)$$

# 特徴の分布の仮定②

## ■ 特徴の確率密度関数

- さらに,  $p(x_i|8)$  と  $p(x_i|B)$  が平均\* $m_i^8$ と $m_i^B$ , 標準偏差 $\sigma$ の**正規分布**に従うと仮定する( $i=1,2$ )

$$p(x_i | 8) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - m_i^8)^2}{2\sigma^2}\right]$$

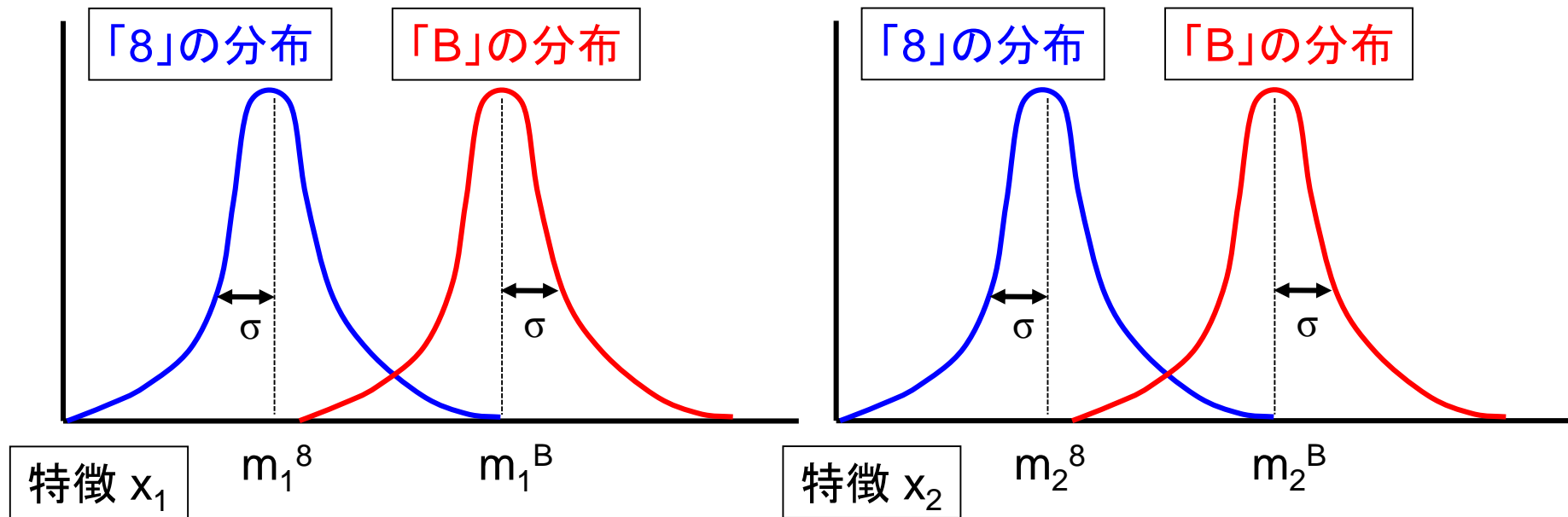
$$m_i^8 = \frac{1}{n_8} \sum_{\mathbf{x} \in \chi_8} x_i$$

$$p(\mathbf{x} | 8) = p(x_1 | 8)p(x_2 | 8) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{m}_8\|^2\right)$$

\*確率密度関数を仮定した場合, パラメータ(平均, 標準偏差)を推定しなければなりません  
(最尤推定にて説明)

# 特徴の分布の仮定②

## 正規分布



# 分散が同じという仮定での予測方法

## ■ 予測方法

□  $p(\mathbf{x}|8) > p(\mathbf{x}|B) \rightarrow$  「8」と予測

$$\frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{m}_8\|^2\right) > \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{m}_B\|^2\right)$$

➡  $\|\mathbf{x} - \mathbf{m}_8\|^2 < \|\mathbf{x} - \mathbf{m}_B\|^2$

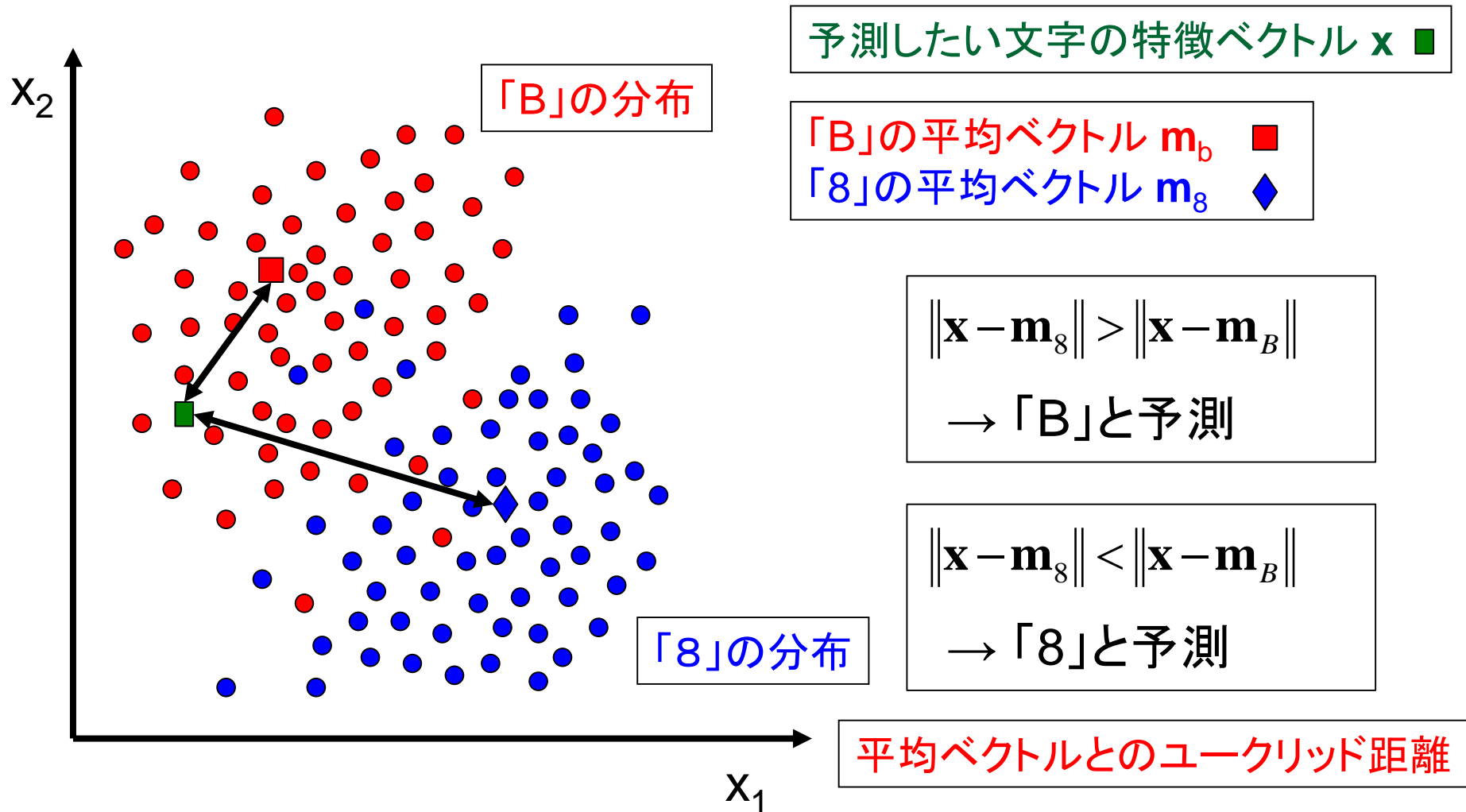
□  $p(\mathbf{x}|8) < p(\mathbf{x}|B) \rightarrow$  「B」と予測

$$\frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{m}_8\|^2\right) < \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{m}_B\|^2\right)$$

➡  $\|\mathbf{x} - \mathbf{m}_8\|^2 > \|\mathbf{x} - \mathbf{m}_B\|^2$

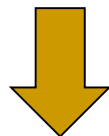


# 平均ベクトルによる予測方法①



## 平均ベクトルによる予測方法②

- 分散が同じという仮定での予測方法



- 各クラスのプロトタイプの特徴ベクトルを平均ベクトルとした最近傍法と等価

# 特徴の分布の仮定③

## ■ 特徴の確率密度関数

- 「8」の特徴ベクトル  $\mathbf{x}$  は, 平均 $\mathbf{m}_8$ , 分散共分散行列 $\Sigma_8$ の正規分布に従う

$$\Sigma_8 = \frac{1}{n_8} \sum_{\mathbf{x} \in \chi_8} (\mathbf{x} - \mathbf{m}_8)(\mathbf{x} - \mathbf{m}_8)^t$$

$n_8$ :「8」のパターン数

$$p(\mathbf{x} | 8) = \frac{1}{2\pi |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8)\right)$$

- 「B」の特徴ベクトル  $\mathbf{x}$  は, 平均 $\mathbf{m}_B$ , 分散共分散行列 $\Sigma_B$ の正規分布に従う

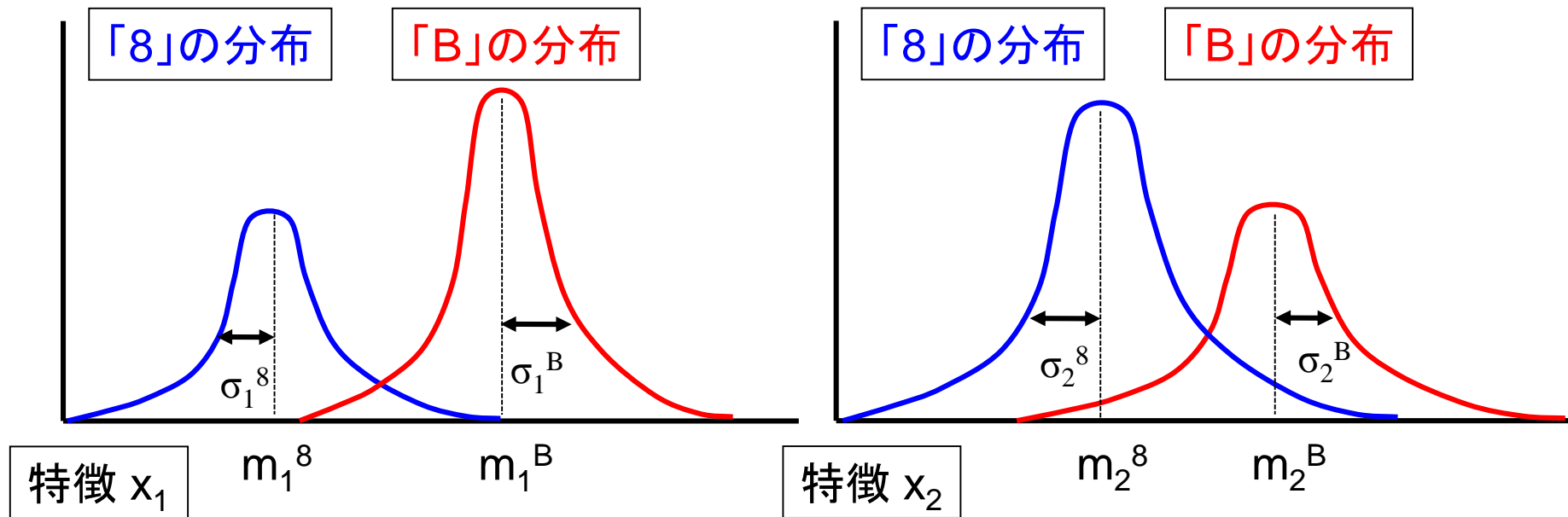
$$\Sigma_B = \frac{1}{n_B} \sum_{\mathbf{x} \in \chi_B} (\mathbf{x} - \mathbf{m}_B)(\mathbf{x} - \mathbf{m}_B)^t$$

$n_B$ :「B」のパターン数

$$p(\mathbf{x} | B) = \frac{1}{2\pi |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B)\right)$$

# 特徴の分布の仮定④

正規分布



# パラメータの推定

- 分布を仮定した場合，パラメータも未知なため推定しなければならない

平均ベクトル

$$\mathbf{m}^g = \frac{1}{n_g} \sum_{\mathbf{x} \in \chi_g} \mathbf{x}$$

分散共分散行列

$$\Sigma_g = \frac{1}{n_g} \sum_{\mathbf{x} \in \chi_g} (\mathbf{x} - \mathbf{m}_g)(\mathbf{x} - \mathbf{m}_g)^t$$

# 正規分布に従うという仮定での予測方法①

## ■ 予測方法

□  $p(\mathbf{x}|8) > p(\mathbf{x}|B) \rightarrow$  「8」と予測

$$\frac{1}{2\pi |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x} - \mathbf{m}_8)\right) > \frac{1}{2\pi |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x} - \mathbf{m}_B)\right)$$

両辺、対数をとると

$$-\frac{1}{2}(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x} - \mathbf{m}_8) - \log 2\pi - \frac{1}{2} \log |\Sigma_8| > -\frac{1}{2}(\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x} - \mathbf{m}_B) - \log 2\pi - \frac{1}{2} \log |\Sigma_B|$$



$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x} - \mathbf{m}_8) + \log |\Sigma_8| < (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x} - \mathbf{m}_B) + \log |\Sigma_B|$$

# 正規分布に従うという仮定での予測方法②

## ■ 予測方法

□  $p(\mathbf{x}|8) < p(\mathbf{x}|B) \rightarrow$  「B」と予測

$$\frac{1}{2\pi |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8)\right) < \frac{1}{2\pi |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B)\right)$$

両辺、対数をとると

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8) - \log 2\pi - \frac{1}{2} \log |\Sigma_8| < -\frac{1}{2}(\mathbf{x}-\mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B) - \log 2\pi - \frac{1}{2} \log |\Sigma_B|$$



$$(\mathbf{x}-\mathbf{m}_8)^t \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8) + \log |\Sigma_8| > (\mathbf{x}-\mathbf{m}_B)^t \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B) + \log |\Sigma_B|$$

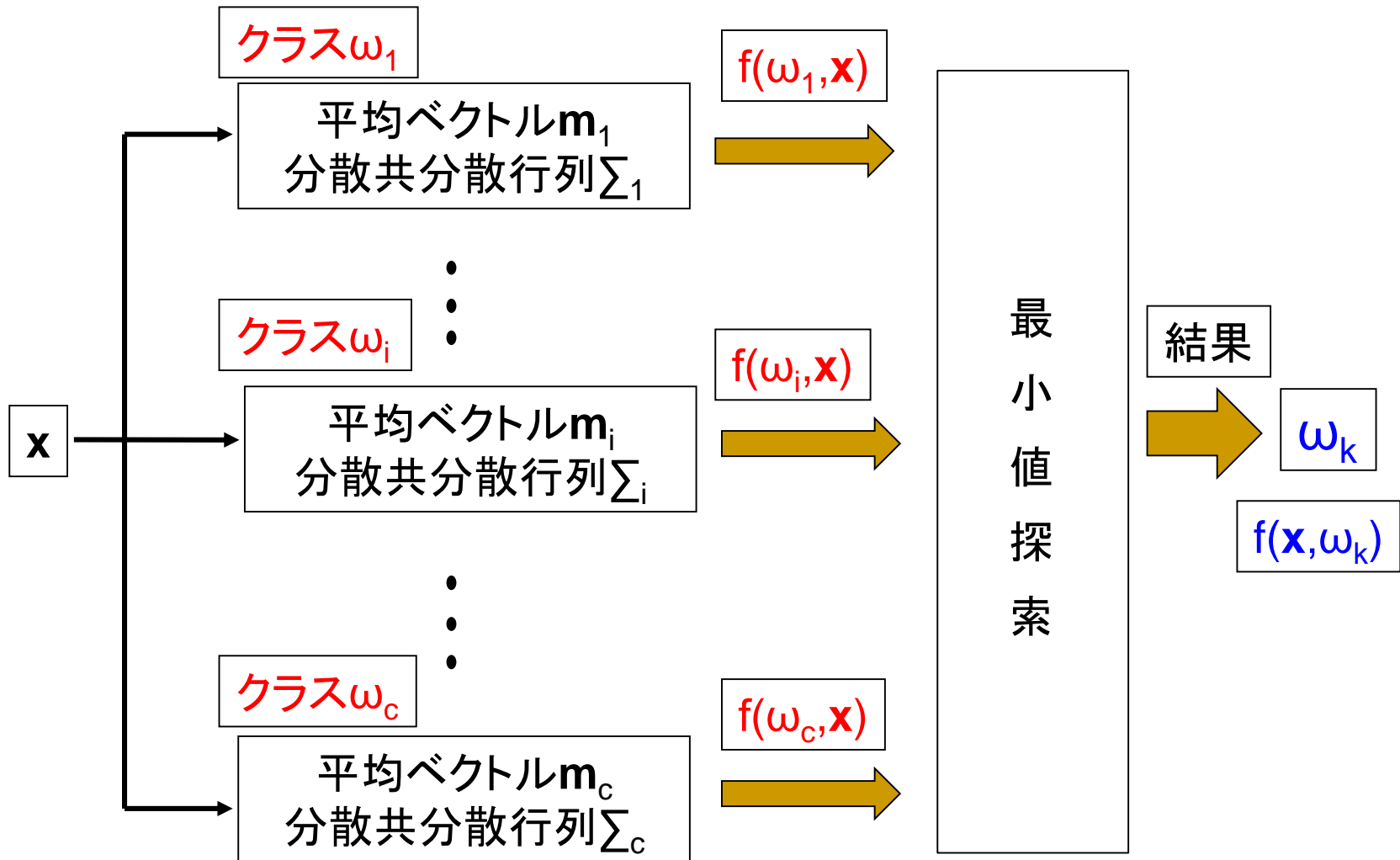
## 正規分布に従うという仮定での予測方法③

- $c$ 個のクラス  $\omega_i$  ( $i=1,2,\dots,c$ )
  - クラス  $\omega_i$  の平均ベクトル  $\mathbf{m}_i$  ( $d$ 次元)
  - 分散共分散行列  $\Sigma_i$
- 予測したいデータの特徴ベクトル  $\mathbf{x}$ 
  - 特徴ベクトル  $\mathbf{x}$  はどのクラスに属するか

$$\begin{aligned} & \min_{i=1,2,\dots,c} (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i) + \log |\Sigma_i| \\ & = (\mathbf{x} - \mathbf{m}_k)^t \Sigma_k^{-1} (\mathbf{x} - \mathbf{m}_k) + \log |\Sigma_k| \Rightarrow \mathbf{x} \in \omega_k \end{aligned}$$



# 正規分布に従うという仮定での予測方法④



$$f(\omega_i, \mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i) + \log |\Sigma_i|$$

# マハラノビス距離による予測方法①

## ■ 予測方法

マハラノビス距離

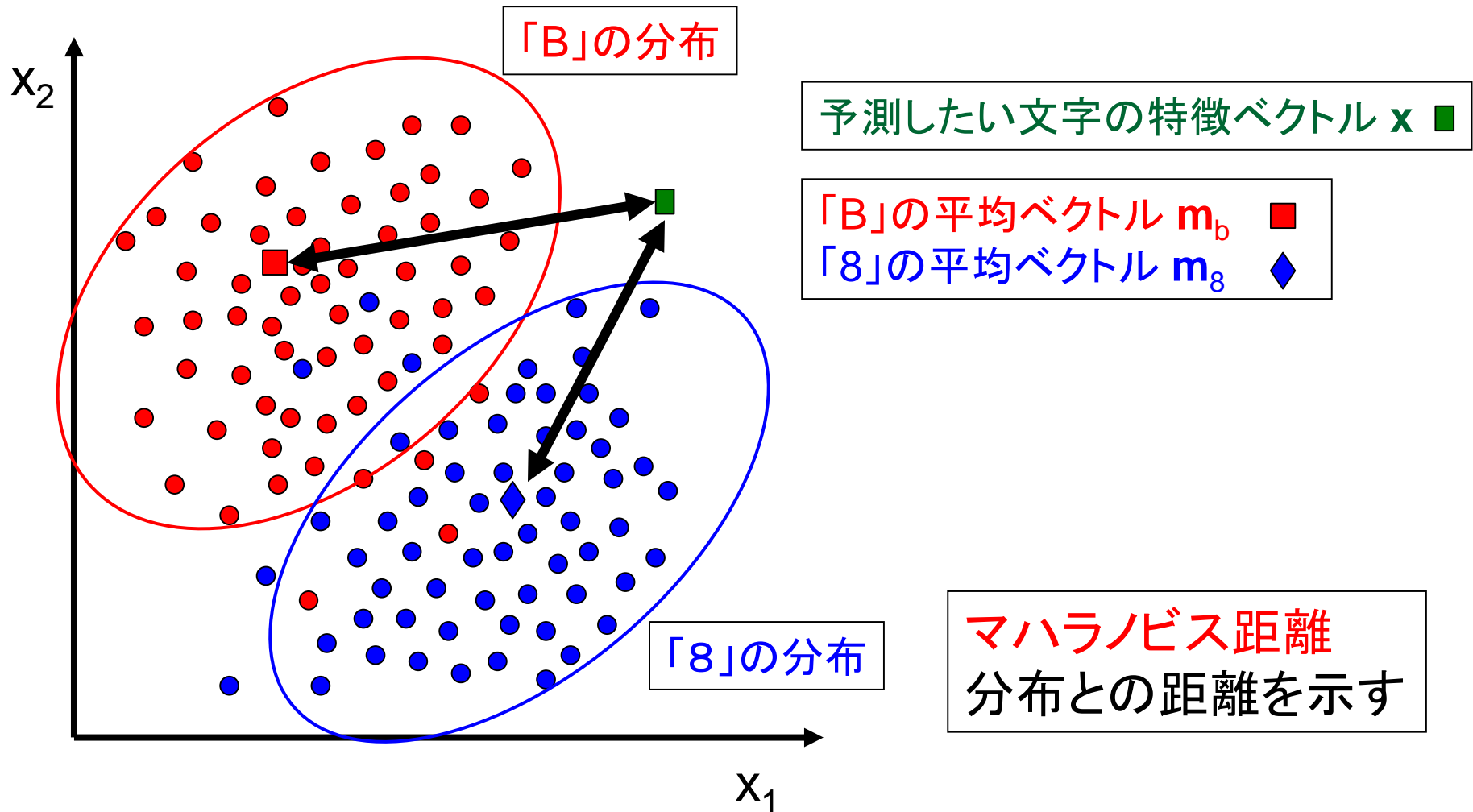
$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) < (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B)$$

➡ 「8」と予測

$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) > (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B)$$

➡ 「B」と予測

# マハラノビス距離による予測方法②



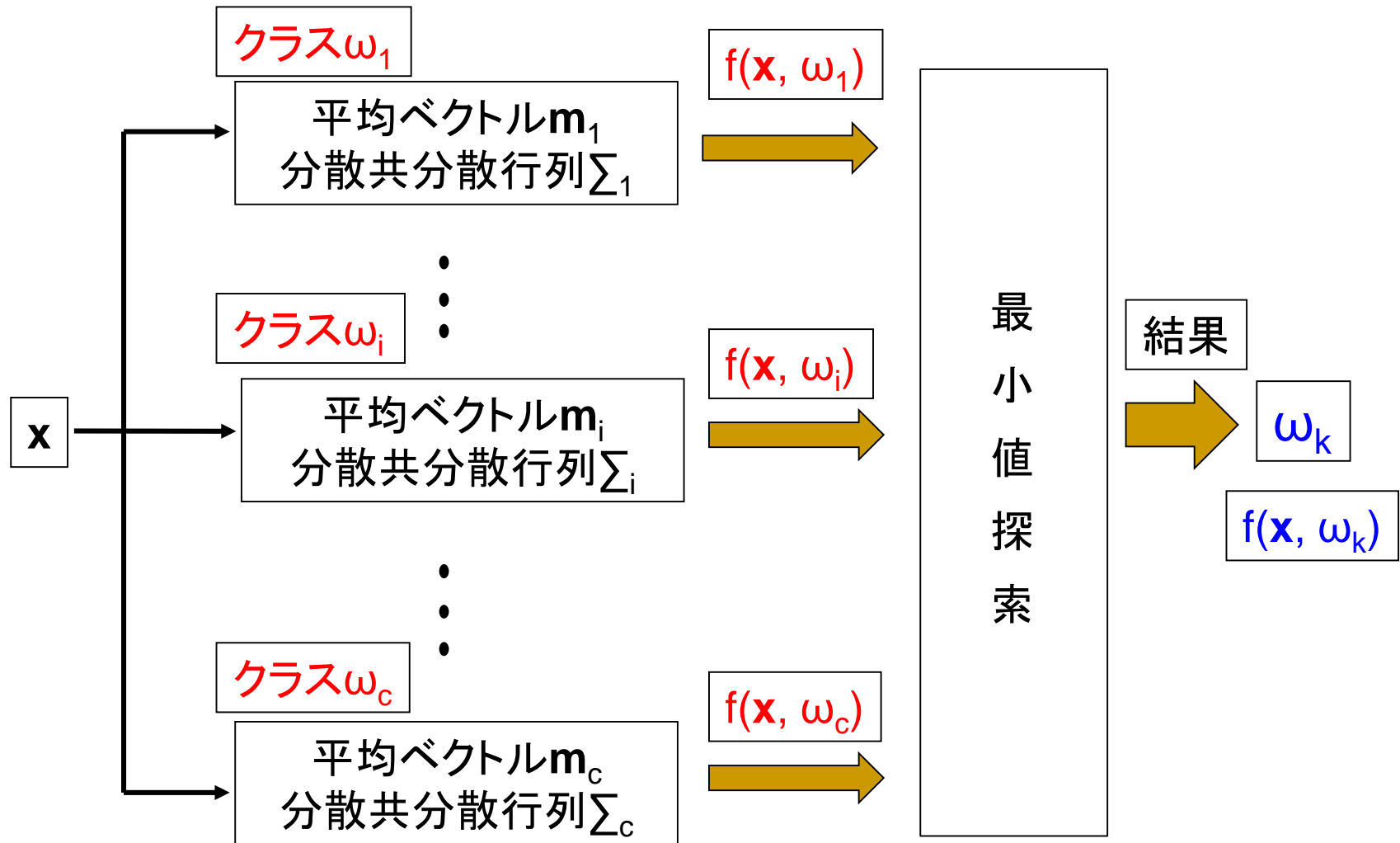
# マハラノビス距離による予測方法③

- $c$ 個のクラス  $\omega_i (i=1,2,\dots,c)$ 
  - クラス  $\omega_i$  の平均ベクトル  $\mathbf{m}_i$  ( $d$ 次元)
  - 分散共分散行列  $\Sigma_i$
- 予測したいデータの特徴ベクトル  $\mathbf{x}$ 
  - 特徴ベクトル  $\mathbf{x}$  はどのクラスに属するか

$$\min_{i=1,2,\dots,c} (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i) = (\mathbf{x} - \mathbf{m}_k)^t \Sigma_k^{-1} (\mathbf{x} - \mathbf{m}_k) \Rightarrow \mathbf{x} \in \omega_k$$

分布とのマハラノビス距離が最小となるクラス

# マハラノビス距離による予測方法④



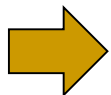
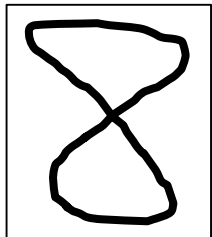
マハラノビス距離

$$f(\mathbf{x}, \omega_i) = (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i)$$

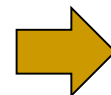
# ベイズ決定則 (統計的機械学習③)

ベイズの定理

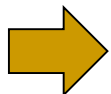
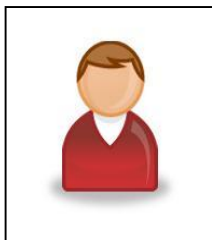
# 予測の手順①



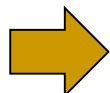
特徴は...  
○が上下に二つ並んでいる



数字の「8」と  
予測



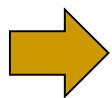
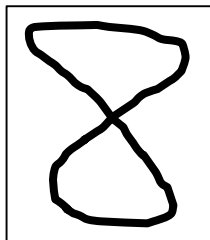
特徴は...  
髪が短い



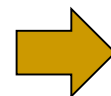
「男性」と予測

特徴を観測後、予測を行なう

## 予測の手順②



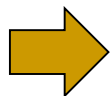
特徴は...  
○が上下に二つ並んでいる



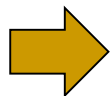
数字の「8」と  
予測

「○が上下に二つ並んでいる場合、8である確率」を基に予測

「○が上下に二つ並んでいる場合、8である確率」  
> 「○が上下に二つ並んでいる場合、Bである確率」



特徴は...  
髪が短い



「男性」と認識

「髪が短い場合、男性である確率」を基に予測

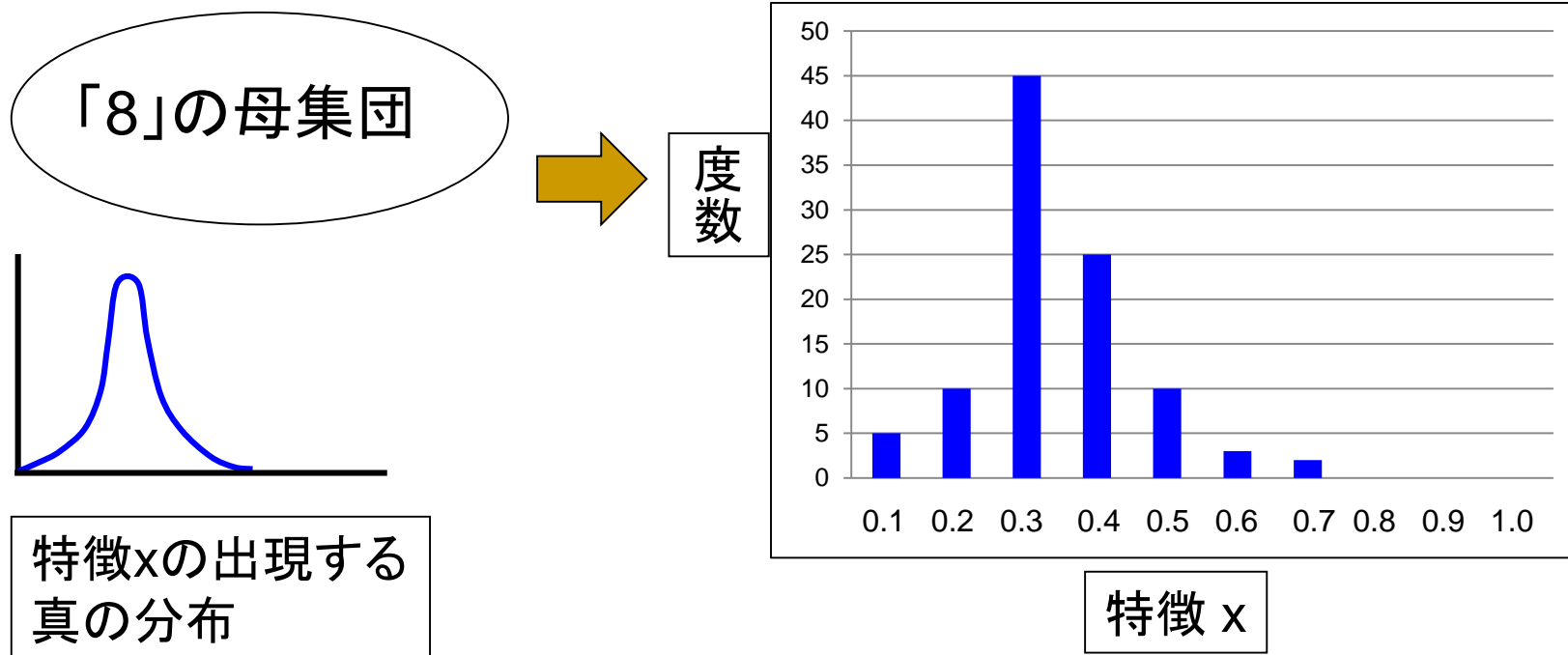
「髪が短い場合、男性である確率」  
> 「髪が短い場合、女性である確率」



# 統計的機械学習①で利用した確率密度関数

## ■ $p(x|8)$

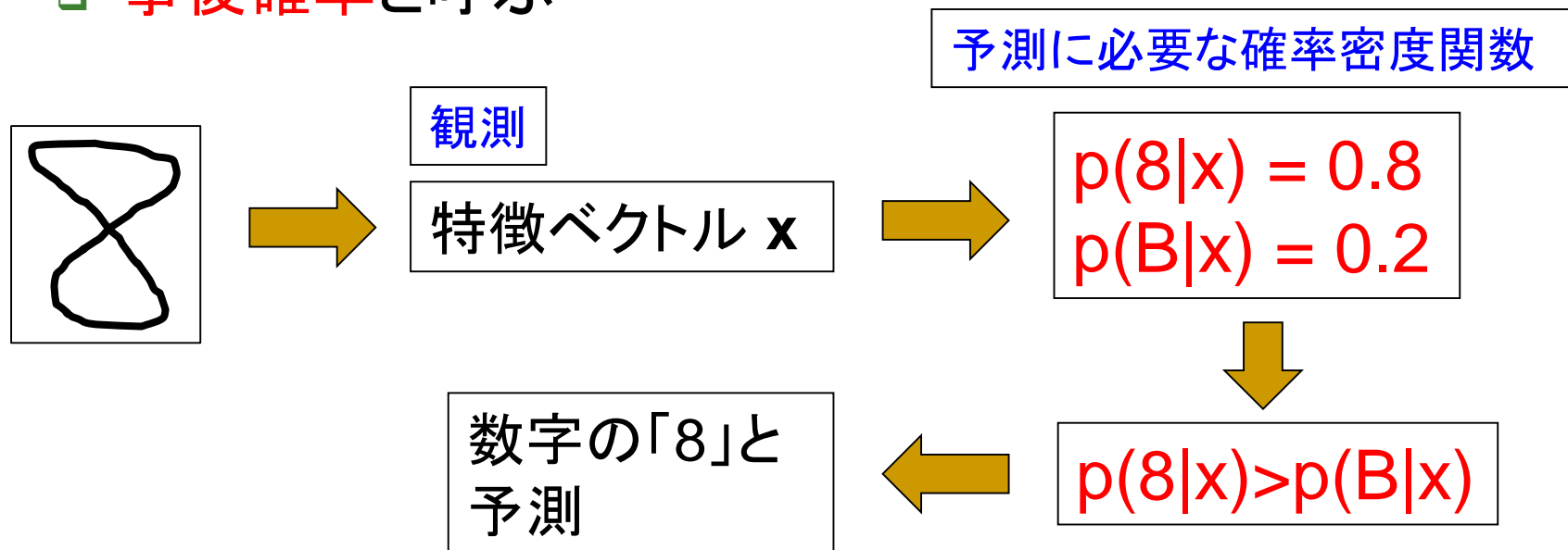
- 「8」というパターンが与えられた場合、特徴 $x$ が出現する確率(条件つき確率)



# 「予測」に必要な確率密度関数

## ■ $p(8|x)$

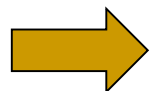
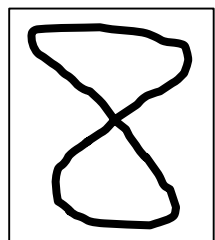
- $x$  という特徴が出現した場合, 「8」が出現する確率
- 事後確率と呼ぶ



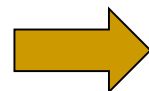
# ベイズ決定則

- $p(8|\mathbf{x})$ ,  $p(B|\mathbf{x})$

- 特徴ベクトル  $\mathbf{x}$  を観測した後の生起確率(事後確率)



特徴ベクトル  $\mathbf{x}$



$p(8|\mathbf{x})$ ,  $p(B|\mathbf{x})$  を求める

$p(8|\mathbf{x}) > p(B|\mathbf{x})$  の場合  
→ 「8」と予測

$p(8|\mathbf{x}) < p(B|\mathbf{x})$  の場合  
→ 「B」と予測

事後確率が最大になるクラスを結果とする(ベイズ決定則)

# 利用する確率

## ■ (例) 男女の判定

条件付き確率

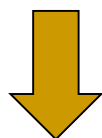
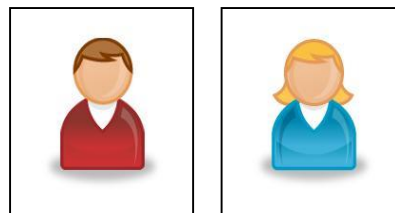
$$p(\text{髪が長い}|\text{男性}) = 0.15$$

$$p(\text{髪が長い}|\text{女性}) = 0.75$$

## ■ 事前確率

- $p(\text{男性}) = 0.5$

- $p(\text{女性}) = 0.5$



髪が長いかどうかという特徴を  
観測すれば...

## ■ 事後確率

- $p(\text{男性}|\text{髪が長い}) = 0.2$

- $p(\text{女性}|\text{髪が長い}) = 0.8$

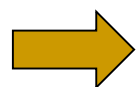
# 事後確率の求め方

## ■ ベイズの定理

$$p(8 | \mathbf{x}) = \frac{p(\mathbf{x} | 8)p(8)}{p(\mathbf{x})}$$
$$p(B | \mathbf{x}) = \frac{p(\mathbf{x} | B)p(B)}{p(\mathbf{x})}$$

$p(\mathbf{x})$  は同じなので,

$$p(\mathbf{x} | 8)p(8) > p(\mathbf{x} | B)p(B)$$

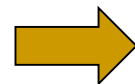


「8」と予測

$p(8)$ ,  $p(B)$   
事前確率と呼ぶ

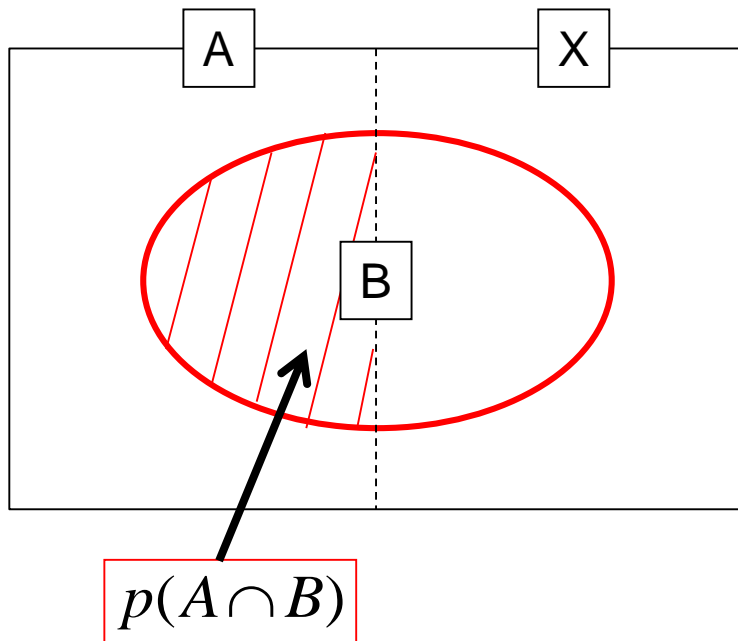
$$p(8 | \mathbf{x}) + p(B | \mathbf{x}) = 1$$
$$p(8) + p(B) = 1$$
$$p(\mathbf{x}) = p(\mathbf{x} | 8)p(8) + p(\mathbf{x} | B)p(B)$$

$$p(\mathbf{x} | 8)p(8) < p(\mathbf{x} | B)p(B)$$



「B」と予測

# ベイズの定理



$$p(A \cap B) = p(A)p(B | A) \\ = p(B)p(A | B)$$



$$p(B | A) = \frac{p(B)p(A | B)}{p(A)}$$

# ベイズ決定則による予測方法①

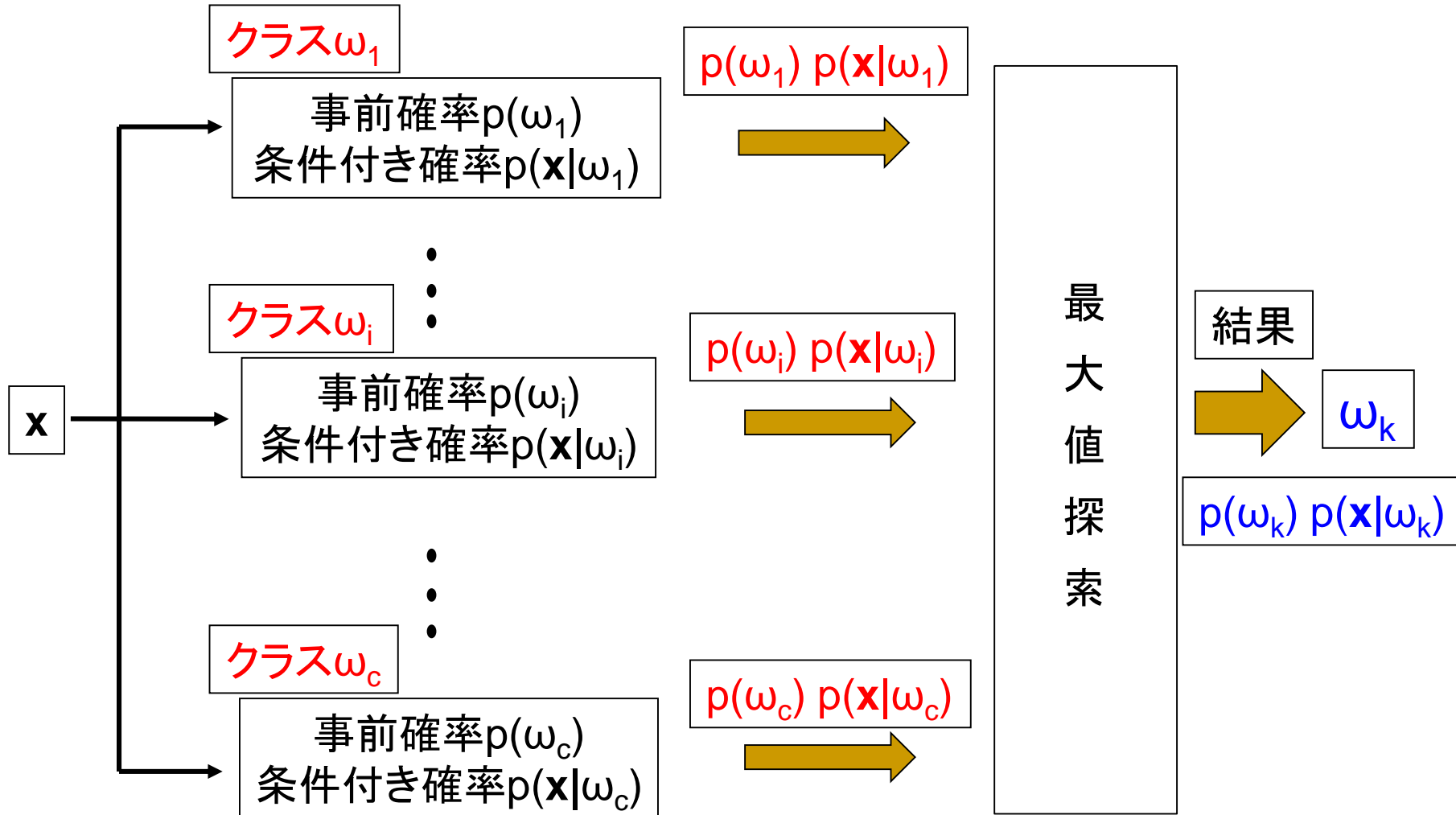
$$p(\omega_i) = \frac{n_i}{\sum_{i=1}^c n_i}$$

$n_i$  クラス $\omega_i$ のデータ数

- $c$ 個のクラス $\omega_i$  ( $i=1, 2, \dots, c$ )
  - 事前確率  $p(\omega_i)$  は既知
- 予測したいデータの特徴ベクトル  $\mathbf{x}$  ( $d$ 次元)
  - 各クラスごとの条件付き確率  $p(\mathbf{x} | \omega_i)$
  - 特徴ベクトル  $\mathbf{x}$  はどのクラスに属するか

$$\begin{aligned} \max_{i=1,2,\dots,c} \{p(\omega_i | \mathbf{x})\} &= \max_{i=1,2,\dots,c} \left\{ \frac{p(\mathbf{x} | \omega_i) p(\omega_i)}{p(\mathbf{x})} \right\} \\ &= \max_{i=1,2,\dots,c} \{p(\mathbf{x} | \omega_i) p(\omega_i)\} \\ &= p(\omega_k | \mathbf{x}) \Rightarrow \mathbf{x} \in \omega_k \end{aligned}$$

# ベイズ決定則による予測方法②





# 数値例(ベイズ決定則)①

文字	事前確率
A	0.2
B	0.15
C	0.25
D	0.05
E	0.35

## 事前確率

左表は5文字の出現確率を示す



特徴を考慮せず予測した場合



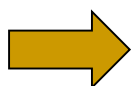
事前確率が最大となる文字を結果とする  
→  $p(E)=0.35$  → 常に「E」と認識

# 数値例(ベイズ決定則)②

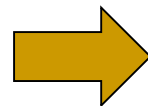
特徴  $x$  の出現確率  $p(x|\text{文字})$   
(特徴  $x$  は 1, 2, 3 という値)

文字	事前確率	特徴 $x$	1	2	3
A	0.2	A	0.5	0.4	0.1
B	0.15	B	0.1	0.8	0.1
C	0.25	C	0.3	0.4	0.3
D	0.05	D	0.1	0.2	0.7
E	0.35	E	0.8	0.1	0.1

C



特徴  $x$  は 3



出現確率が最大となるのは  
 $P(x=3|D)=0.7 \rightarrow$  結果は「D」

本当？

# 数値例(ベイズ決定則)③

## ■ 事後確率の計算

文字	事前確率	特徴x	1	2	3
A	0.2	A	0.5	0.4	0.1
B	0.15	B	0.1	0.8	0.1
C	0.25	C	0.3	0.4	0.3
D	0.05	D	0.1	0.2	0.7
E	0.35	E	0.8	0.1	0.1

x=3の場合の「A」の出現確率

$$P(A|x=3) = \frac{p(x=3|A)p(A)}{p(x=3)} = \frac{0.1 \times 0.2}{0.1 \times 0.2 + 0.1 \times 0.15 + 0.3 \times 0.25 + 0.7 \times 0.05 + 0.1 \times 0.35} = \frac{0.02}{0.18}$$

# 数値例(ベイズ決定則)④

## ■ 事後確率の計算

$$P(B|x=3) = \frac{p(x=3|B)p(B)}{p(x=3)} = \frac{0.15 \times 0.1}{0.1 \times 0.2 + 0.1 \times 0.15 + 0.3 \times 0.25 + 0.7 \times 0.05 + 0.1 \times 0.35} = \frac{0.015}{0.18}$$

$$P(C|x=3) = \frac{p(x=3|C)p(C)}{p(x=3)} = \frac{0.25 \times 0.3}{0.1 \times 0.2 + 0.1 \times 0.15 + 0.3 \times 0.25 + 0.7 \times 0.05 + 0.1 \times 0.35} = \frac{0.075}{0.18}$$

$$P(D|x=3) = \frac{p(x=3|D)p(D)}{p(x=3)} = \frac{0.05 \times 0.7}{0.1 \times 0.2 + 0.1 \times 0.15 + 0.3 \times 0.25 + 0.7 \times 0.05 + 0.1 \times 0.35} = \frac{0.035}{0.18}$$

$$P(E|x=3) = \frac{p(x=3|E)p(E)}{p(x=3)} = \frac{0.35 \times 0.1}{0.1 \times 0.2 + 0.1 \times 0.15 + 0.3 \times 0.25 + 0.7 \times 0.05 + 0.1 \times 0.35} = \frac{0.035}{0.18}$$

事後確率が最大となるは「C」→「C」と予測

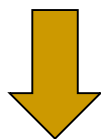
# 事後確率の計算

- 事後確率を計算するには？

- 事前確率  $p(\omega_i)$  は既知

$$p(\omega_i) = \frac{n_i}{\sum_{i=1}^c n_i}$$

$n_i$  クラス $\omega_i$ のデータ数



- 条件付き確率  $p(\mathbf{x}|\omega_i)$  を求める必要がある

- 確率密度関数は仮定する
- パラメータの推定が必要

# 確率密度関数の仮定

## ■ 特徴の確率密度関数

- 「8」の特徴ベクトル  $\mathbf{x}$  (d次元) は, **独立\***でかつ平均 $\mathbf{m}_8$ , 分散共分散行列 $\Sigma_8$ の**正規分布**に従う

$$p(\mathbf{x} | 8) = p(x_1, x_2, \dots, x_d | 8) = p(x_1 | 8) p(x_2 | 8) \cdots p(x_d | 8)$$
$$= \frac{1}{(2\pi)^{d/2} |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8)\right)$$

d次元の正規分布

- 「B」の特徴ベクトル  $\mathbf{x}$  は, **独立**でかつ平均 $\mathbf{m}_B$ , 分散共分散行列 $\Sigma_B$ の**正規分布**に従う

$$p(\mathbf{x} | B) = \frac{1}{(2\pi)^{d/2} |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B)\right)$$

\*独立性を仮定したベイズ決定則を**単純ベイズ**(**ナীবベイズ**)と呼ぶ

# 正規分布の仮定でのベイズ決定則①

- $p(\mathbf{x}|8)p(8) > p(\mathbf{x}|B)p(B) \rightarrow \text{「8」と予測}$

$n_8$ :「8」のパターン数

$n_B$ :「B」のパターン数

$$\underbrace{\frac{1}{(2\pi)^{d/2} |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8)\right)}_{p(\mathbf{x}|8)} \times \underbrace{\frac{n_8}{n_8 + n_B}}_{p(8)} > \underbrace{\frac{1}{(2\pi)^{d/2} |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B)\right)}_{p(\mathbf{x}|B)} \times \underbrace{\frac{n_B}{n_8 + n_B}}_{p(B)}$$

両辺, 対数をとると

$$-\frac{1}{2}(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_8| + \log\left(\frac{n_8}{n_8 + n_B}\right) > -\frac{1}{2}(\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_B| + \log\left(\frac{n_B}{n_8 + n_B}\right)$$



$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) + \log |\Sigma_8| - 2 \log\left(\frac{n_8}{n_8 + n_B}\right) < (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B) + \log |\Sigma_B| - 2 \log\left(\frac{n_B}{n_8 + n_B}\right)$$

# 正規分布の仮定でのベイズ決定則②

■  $p(\mathbf{x}|8)p(8) < p(\mathbf{x}|B)p(B) \rightarrow \text{「B」と予測}$

$$\underbrace{\frac{1}{(2\pi)^{d/2} |\Sigma_8|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_8)' \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8)\right)}_{p(\mathbf{x}|8)} \times \underbrace{\frac{n_8}{n_8+n_B}}_{p(8)} < \underbrace{\frac{1}{(2\pi)^{d/2} |\Sigma_B|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_B)' \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B)\right)}_{p(\mathbf{x}|B)} \times \underbrace{\frac{n_B}{n_8+n_B}}_{p(B)}$$

両辺、対数をとると

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_8)' \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_8| + \log\left(\frac{n_8}{n_8+n_B}\right) < -\frac{1}{2}(\mathbf{x}-\mathbf{m}_B)' \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_B| + \log\left(\frac{n_B}{n_8+n_B}\right)$$

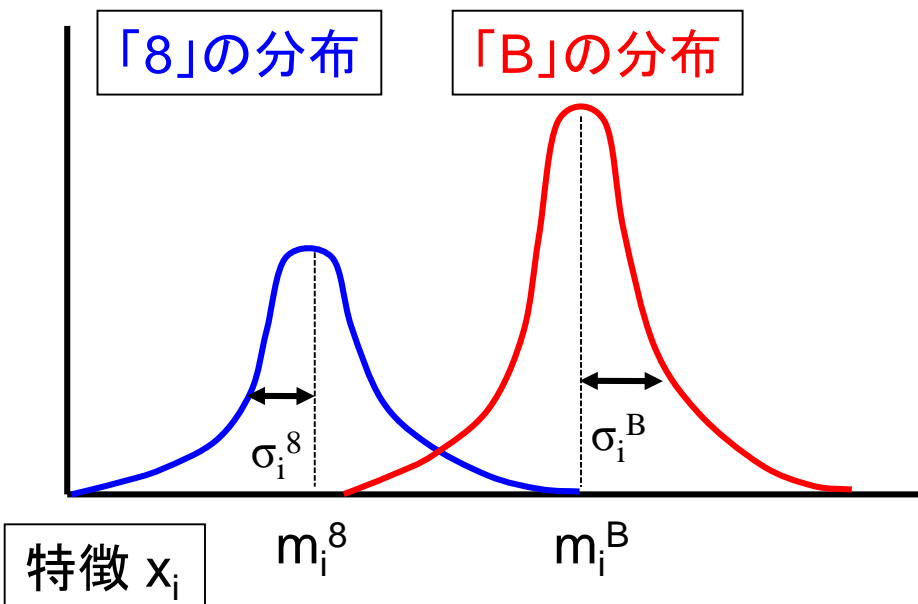


$$(\mathbf{x}-\mathbf{m}_8)' \Sigma_8^{-1}(\mathbf{x}-\mathbf{m}_8) + \log |\Sigma_8| - 2 \log\left(\frac{n_8}{n_8+n_B}\right) > (\mathbf{x}-\mathbf{m}_B)' \Sigma_B^{-1}(\mathbf{x}-\mathbf{m}_B) + \log |\Sigma_B| - 2 \log\left(\frac{n_B}{n_8+n_B}\right)$$



# 正規分布を仮定した場合のベイズ決定則①

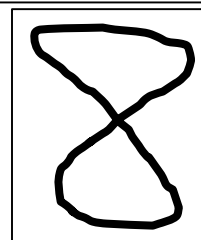
条件付き確率  $p(x_i|8)$   $p(x_i|B)$



正規分布を仮定  
平均ベクトル ( $\mathbf{m}_8, \mathbf{m}_B$ )  
分散共分散行列 ( $\Sigma_8, \Sigma_B$ ) を推定

事前確率  $p(8)$   $p(B)$  は既知

未知文字を予測したい場合



特徴ベクトル  $\mathbf{x}$

事後確率を計算  
 $p(8|\mathbf{x})$   $p(B|\mathbf{x})$

## 正規分布を仮定した場合のベイズ決定則②

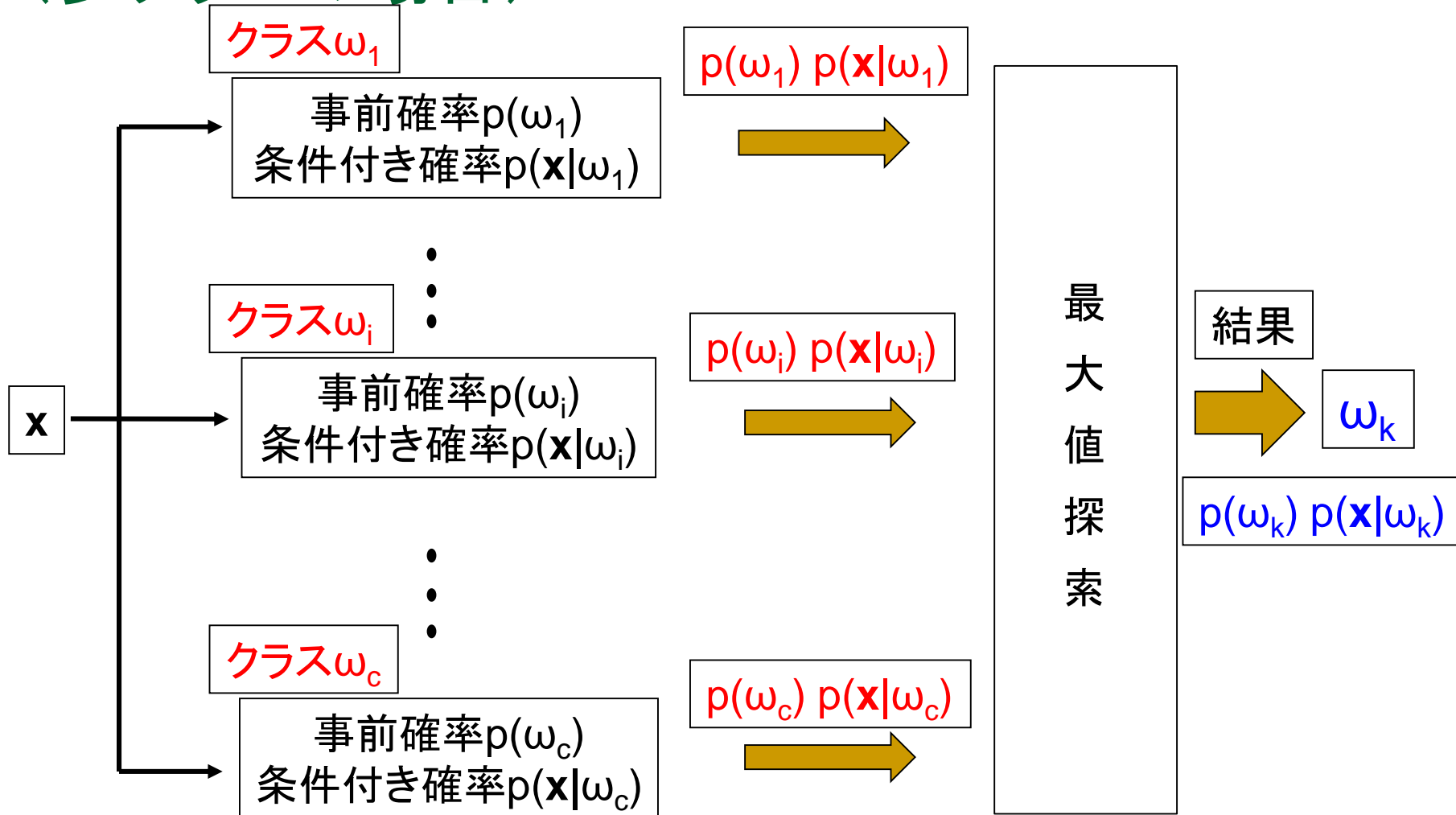
- $p(\mathbf{x}|8)p(8) > p(\mathbf{x}|B)p(B) \rightarrow$  「8」と予測

$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) + \log |\Sigma_8| - 2 \log \left( \frac{n_8}{n_8 + n_B} \right) < (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B) + \log |\Sigma_B| - 2 \log \left( \frac{n_B}{n_8 + n_B} \right)$$

- $p(\mathbf{x}|8)p(8) < p(\mathbf{x}|B)p(B) \rightarrow$  「B」と予測

$$(\mathbf{x} - \mathbf{m}_8)^t \Sigma_8^{-1} (\mathbf{x} - \mathbf{m}_8) + \log |\Sigma_8| - 2 \log \left( \frac{n_8}{n_8 + n_B} \right) > (\mathbf{x} - \mathbf{m}_B)^t \Sigma_B^{-1} (\mathbf{x} - \mathbf{m}_B) + \log |\Sigma_B| - 2 \log \left( \frac{n_B}{n_8 + n_B} \right)$$

# 正規分布を仮定した場合のベイズ決定則 (多クラスの場合)



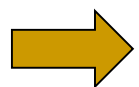
$$p(\mathbf{x} | \omega_i) = -\frac{1}{2} (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i) - \frac{1}{2} \log |\Sigma_i| + \log p(\omega_i)$$

# 特徴の出現確率を用いた予測とは？①

## (統計的機械学習①とは？)

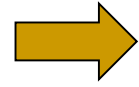
### ■ ベイズ決定則

$$p(8 | \mathbf{x}) > p(B | \mathbf{x})$$
$$p(\mathbf{x} | 8)p(8) > p(\mathbf{x} | B)p(B)$$



「8」と予測

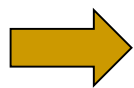
$$p(8 | \mathbf{x}) > p(B | \mathbf{x})$$
$$p(\mathbf{x} | 8)p(8) < p(\mathbf{x} | B)p(B)$$



「B」と予測

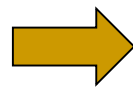
$p(8) = p(B)$  の場合

$$p(\mathbf{x} | 8) > p(\mathbf{x} | B)$$



「8」と予測

$$p(\mathbf{x} | 8) < p(\mathbf{x} | B)$$



「B」と予測

# 特徴の出現確率を用いた認識とは？②

## (統計的機械学習①とは？)

- ベイズ決定則において、事前確率  $p(g)$ ,  $p(B)$  ( $n_g$ と  $n_B$ ) が同じ場合、特徴の出現確率のみを用いて予測が可能
- ただし,
  - 「 $p(x|g)+p(x|B)=1$ 」は成立せず、「確率」の大小によって予測していると言い難い
  - この場合、 $p(x|g)$ ,  $p(x|B)$  を**尤度**と呼び、この手法を**最尤法\***と呼ぶ

\*最尤推定とは違うことに注意

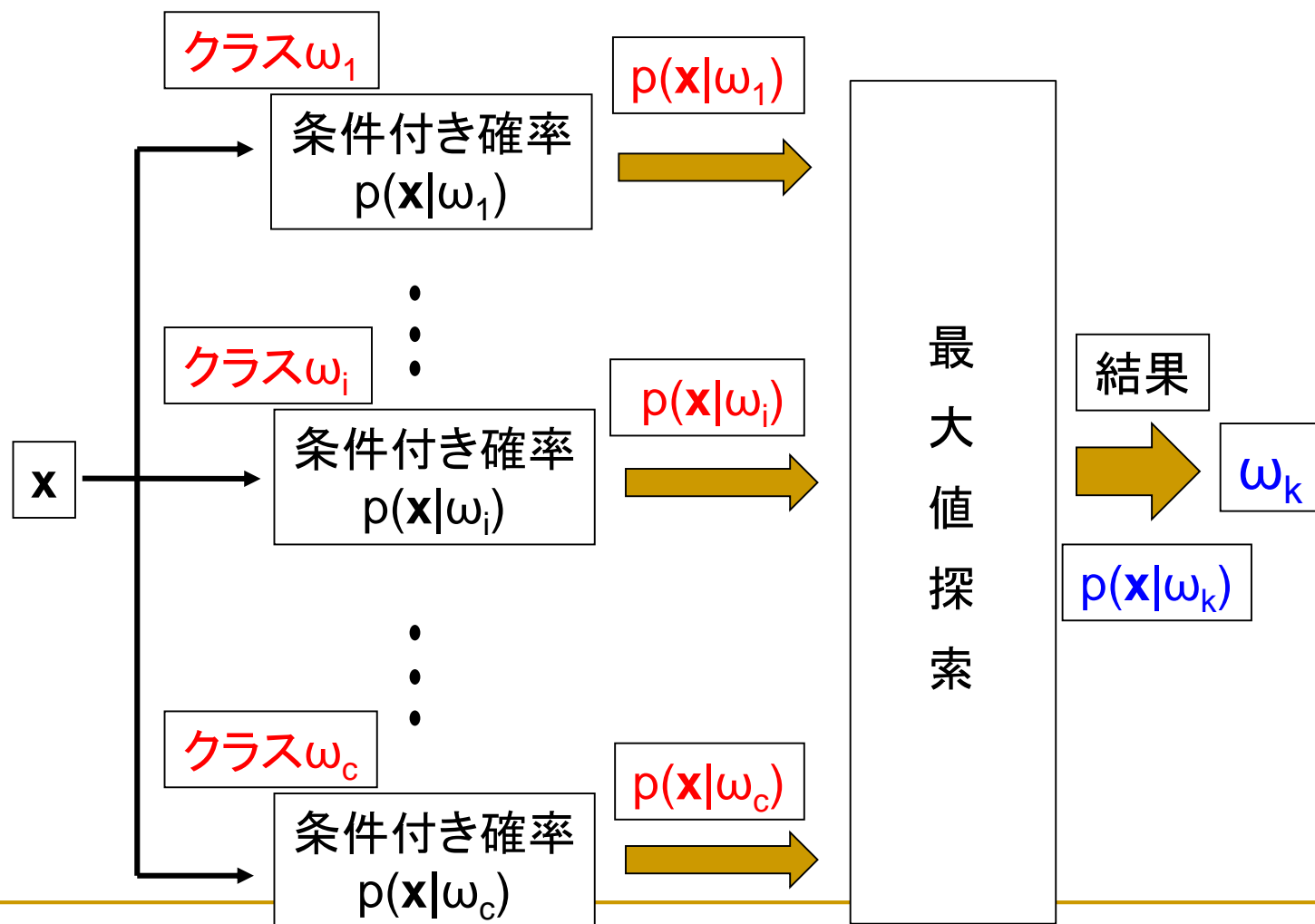
# 最尤法による予測①

## (統計的機械学習①)

- $c$ 個のクラス $\omega_i (i=1, 2, \dots, c)$ 
  - 事前確率  $p(\omega_i)$  は全て同じ ( $p(\omega_i) = 1/c$ )
- 予測したいデータの特徴ベクトル  $\mathbf{x}$  ( $d$ 次元)
  - 各クラスごとの条件付き確率(尤度)  $p(\mathbf{x} | \omega_i)$
  - 特徴ベクトル  $\mathbf{x}$  はどのクラスに属するか

$$\begin{aligned} & \max_{i=1, 2, \dots, c} \{p(\mathbf{x} | \omega_i)\} \\ & = p(\omega_k | \mathbf{x}) \Rightarrow \mathbf{x} \in \omega_k \end{aligned}$$

# 最尤法による認識② (統計的機械学習①)



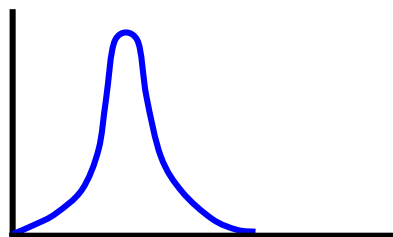
# パラメータの推定

最尤推定

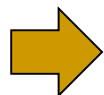


# パラメータの推定方法①

母集団

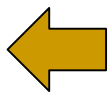
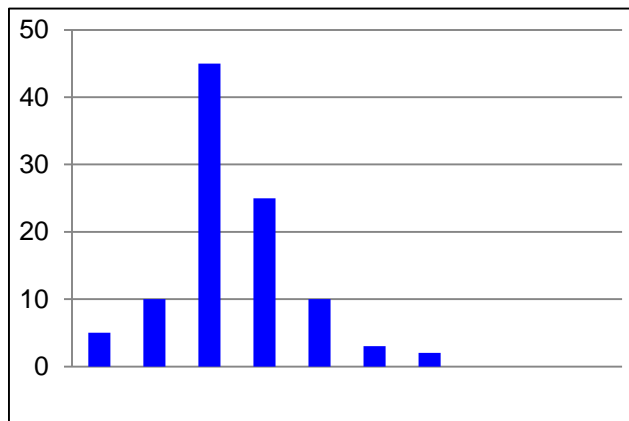


特徴xの出現する  
真の分布



特徴x

0.1, 0.4, 0.3, 0.5 ...

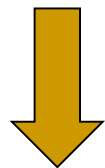


真の分布を推定

- ① 確率密度関数を仮定
- ② パラメータを推定

# パラメータの推定方法②

- 確率密度関数は正規分布と仮定
- パラメータ(平均, 分散共分散行列)はどのような値を用いればよいか



結論としては...

$$\hat{m}_g = \frac{1}{n_g} \sum_{\mathbf{x} \in g} \mathbf{x}$$

$$\hat{\Sigma}_g = \frac{1}{n_g} \sum_{\mathbf{x} \in g} (\mathbf{x} - \hat{\mathbf{m}}_g)(\mathbf{x} - \hat{\mathbf{m}}_g)^t$$

$$\hat{m}_B = \frac{1}{n_B} \sum_{\mathbf{x} \in B} \mathbf{x}$$

$$\hat{\Sigma}_B = \frac{1}{n_B} \sum_{\mathbf{x} \in B} (\mathbf{x} - \hat{\mathbf{m}}_B)(\mathbf{x} - \hat{\mathbf{m}}_B)^t$$

# 尤度関数①

## ■ コイン投げ



- 5回投げた場合

- 観測データ  $\chi$  : 表, 表, 裏, 裏, 表

- $P(\text{表}) = \theta$

- $P(\text{裏}) = 1 - \theta$

- $\theta$ を推定したい(二項分布のパラメータ推定)



$$\hat{\theta} = 0.6$$

# 尤度関数②

## ■ $\theta$ の推定方法

- $\theta$ の元で、観測データ  $\chi$  が得られる確率

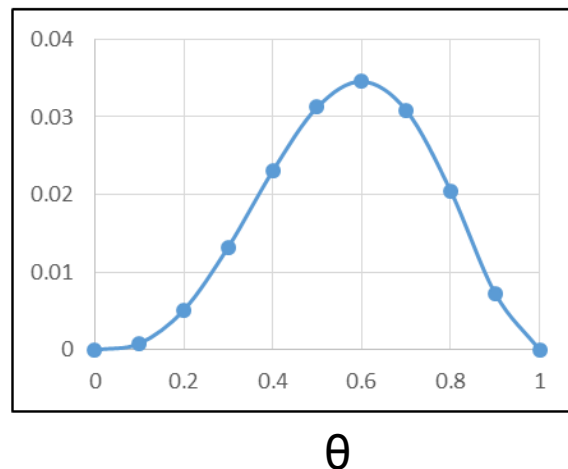
$$L(\chi; \theta) = p(\chi | \theta) = \prod_{k=1}^5 p(x_k | \theta)$$

尤度(関数)

- 観測データ  $\chi$  : 表, 表, 裏, 裏, 表

$$L(\chi; \theta) = \theta \times \theta \times (1 - \theta) \times (1 - \theta) \times \theta$$

$L(\chi; \theta)$



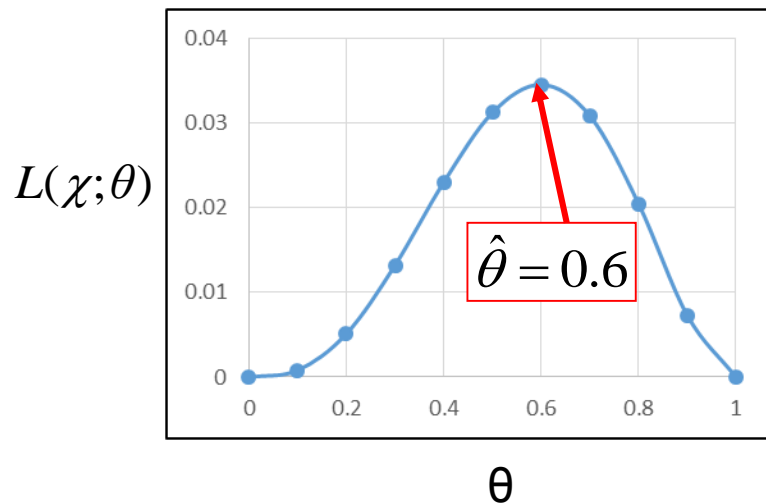
\*尤もらしさ(尤度)と呼ばれています

# 最尤値の求め方①

## ■ $\theta$ の推定方法(最尤推定\*)

- 尤度関数を最大にするパラメータを最尤値とする

$$L(\chi; \theta) = \theta \times \theta \times (1 - \theta) \times (1 - \theta) \times \theta$$



\*参考書によっては最尤法とも呼ぶ場合もあります

# 最尤値の求め方②

- $\chi$ : コインをN回投げたところ, n回表が出た

$$L(\chi; \theta) = p(\chi | \theta) = \binom{N}{n} \theta^n (1 - \theta)^{N-n}$$

最大となる $\theta$ を求める

$$\frac{dL(\chi; \theta)}{d\theta} = \binom{N}{n} (n\theta^{n-1}(1-\theta)^{N-n} - (N-n)\theta^n(1-\theta)^{N-n-1})$$

$$= \binom{N}{n} \theta^{n-1} (1-\theta)^{N-n-1} (n(1-\theta) - (N-n)\theta)$$



$$n(1-\theta) - (N-n)\theta = 0$$

$$n - N\theta = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{n}{N}$$

# 最尤推定\*①

- $\omega_i$  に属する  $n$  個のデータ  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  の集合を  $\chi_i$
- $\theta_i$  は推定したいパラメータ ( $\mathbf{m}_i, \Sigma_i$ )
- 確率密度関数  $p(\mathbf{x} | \omega_i; \theta_i)$
- 各データ  $\mathbf{x}_i$  は確率密度関数に従って独立に生起
- 集合  $\chi_i$  が得られる確率  $p(\chi_i; \theta_i)$

$$p(\chi_i; \theta_i) = \prod_{k=1}^n p(\mathbf{x}_k | \omega_i; \theta_i)$$

尤度(関数)

\*参考書によっては最尤法とも呼ぶ場合もあります

# 最尤推定②

- 尤度を最大にする  $\theta_i$  を求める(最尤推定)

$$\max_{\theta_i} \{ p(\chi_i; \theta_i) \} = p(\chi_i; \hat{\theta}_i) = \prod_{k=1}^n p(\mathbf{x}_k | \omega_i; \hat{\theta}_i)$$

対数をとった場合  
(対数尤度)

$$\frac{\partial}{\partial \theta_i} p(\chi_i; \theta_i) = \mathbf{0}$$

$$\frac{\partial}{\partial \theta_i} \log p(\chi_i; \theta_i)$$

$$= \frac{\partial}{\partial \theta_i} \log \prod_{k=1}^n p(\mathbf{x}_k | \omega_i; \theta_i)$$

$$= \sum_{k=1}^n \frac{\partial}{\partial \theta_i} \log p(\mathbf{x}_k | \omega_i; \theta_i) = \mathbf{0}$$



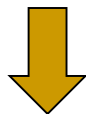
# 最尤推定③

d次元の正規分布

$$p(\mathbf{x}_k | \omega_i; \mathbf{m}_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x}_k - \mathbf{m}_i)\right)$$

$$\sum_{k=1}^n \frac{\partial}{\partial \mathbf{m}_i} \log p(\mathbf{x}_k | \omega_i; \mathbf{m}_i, \Sigma_i) = \mathbf{0}$$

$$\sum_{k=1}^n \frac{\partial}{\partial \Sigma_i} \log p(\mathbf{x}_k | \omega_i; \mathbf{m}_i, \Sigma_i) = \mathbf{0}$$



$$\hat{\mathbf{m}}_i = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

$$\hat{\Sigma}_i = \frac{1}{n} \sum_{i=1}^n (\mathbf{x} - \hat{\mathbf{m}}_i)(\mathbf{x} - \hat{\mathbf{m}}_i)^t$$

平均ベクトル, 分散共分散行列は一般的に求める方法と同じ

# 統計的機械学習のまとめ①

## ■ 統計的機械学習

- 未知のパターンから特徴  $x$  を求め、特徴の出現確率を利用し、どのクラスに属するのかを求めたい

### 尤度

- ① ~~特徴の出現確率を用いた認識~~ (最尤法)
- ② 決定境界による認識
- ③ ベイズ決定則

最尤推定により、確率密度関数のパラメータを求める

# 統計的機械学習のまとめ②

## データの生成

$$P(\mathbf{x}, \omega) = P(\mathbf{x}|\omega)P(\omega)$$

$P(\mathbf{x}, \omega)$  が分かれば無限にデータを生成できる

同時確率

$$P(\mathbf{x}, \omega)$$

母集団

クラス  $\omega$

事前確率

$$P(\omega)$$

条件付き確率

$$P(\mathbf{x}|\omega)$$

## データの特徴

$P(\mathbf{x}|\omega)$  が分かれば特徴の分布が分かる

事後確率

$$P(\omega|\mathbf{x})$$

$P(\omega|\mathbf{x})$  が分かれば予測できる

データの予測

観測データ  $\mathbf{x}$

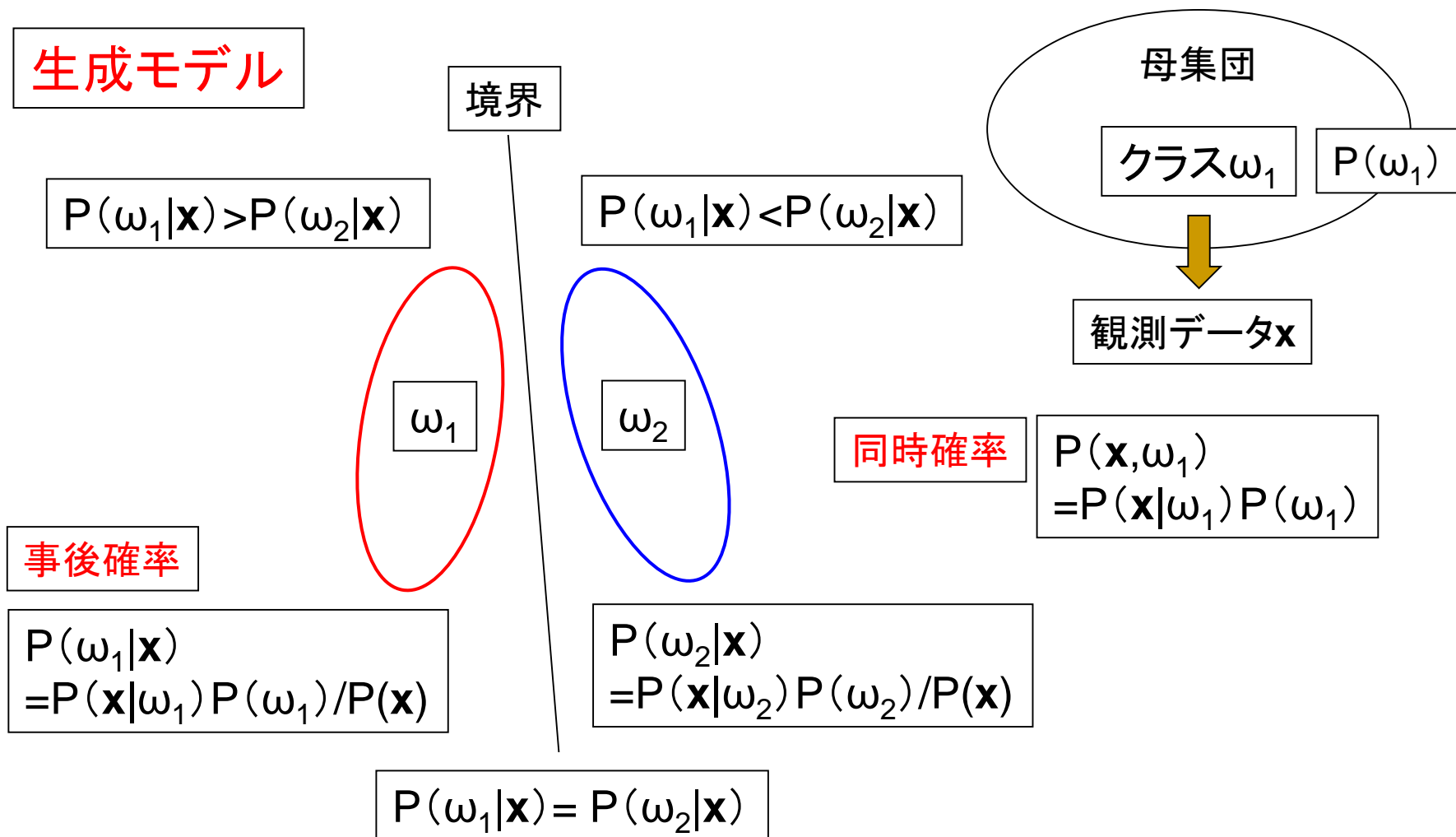


$$P(\omega|\mathbf{x}) = P(\mathbf{x}|\omega)P(\omega)/P(\mathbf{x})$$

同時確率を間接的に用いて予測 (生成モデル)

# 生成モデル

## 生成モデル



# 識別モデル

識別モデル

境界

$$f(x, \omega_1|\theta) = P(\omega_1|x)$$

$\theta$ : パラメータ

$\omega_1$

$\omega_2$

$$f(x, \omega_2|\theta) = P(\omega_2|x)$$

$$f(x, \omega_1|\theta) > f(x, \omega_2|\theta)$$

$$f(x, \omega_1|\theta) < f(x, \omega_2|\theta)$$

$$f(x, \omega_1|\theta) = f(x, \omega_2|\theta)$$

# 実習①

## ベイズ決定則(表計算)

# ベイズ決定則(表計算)

## ■ 「頭痛」の日の予測

## ■ 特徴

- 天気 … 晴れ, 曇り, 雨
- 気温 … 暑い, 適温, 寒い
- 湿度 … 高い, 適当, 低い
- 講義 … yes, no

## ■ 「頭痛」

- 天気=晴れ, 気温=適温, 湿度=適当, 講義=yes
- 頭痛はyes, no の二クラス  
→ 「yes」と「no」のどちらのクラスに属するのか

# 数値データ例

	天気	気温	湿度	講義	頭痛
1	晴れ	寒い	低い	yes	yes
2	曇り	暑い	高い	no	yes
3	晴れ	暑い	低い	no	no
4	雨	適温	高い	yes	no
5	雨	寒い	高い	no	no
6	晴れ	適温	適当	no	yes
7	曇り	暑い	低い	yes	no
8	雨	寒い	高い	no	no
9	曇り	適温	低い	yes	yes
10	曇り	適温	適当	no	no
11	晴れ	暑い	適当	yes	yes
12	晴れ	適温	高い	no	no
13	雨	暑い	低い	no	no
14	雨	寒い	適当	yes	yes
15	曇り	適温	低い	yes	no
16	晴れ	暑い	適当	no	yes
17	晴れ	寒い	高い	yes	yes
18	曇り	適温	高い	yes	no
19	曇り	適温	適当	no	no
20	雨	寒い	適当	no	yes

データ

「統計的」に処理するためには  
多くのデータが必要



この場合、データ数は20個



条件付き確率  $p(\mathbf{x}|\omega_i)$  が求められない場合がある

$p(\text{天気=晴れ, 気温=適温, 湿度=適当, 講義=yes}|\text{頭痛=yes})$  はデータ中に存在しない



# 単純ベイズ①

## ベイズ決定則

- $c$ 個のクラス $\omega_i (i=1,2,\dots,c)$ 
  - 事前確率  $p(\omega_i)$  は既知
- 予測したいデータの特徴ベクトル  $\mathbf{x}$  ( $d$ 次元)
  - 条件付き確率  $p(\mathbf{x} | \omega_i)$
  - 特徴ベクトル  $\mathbf{x}$  はどのクラスに属するか

$$\begin{aligned}\max_{i=1,2,\dots,c} \{ p(\omega_i | \mathbf{x}) \} &= \max_{i=1,2,\dots,c} \left\{ \frac{p(\mathbf{x} | \omega_i) p(\omega_i)}{p(\mathbf{x})} \right\} \\ &= \max_{i=1,2,\dots,c} \{ p(\mathbf{x} | \omega_i) p(\omega_i) \} \\ &= p(\omega_k | \mathbf{x}) \Rightarrow \mathbf{x} \in \omega_k\end{aligned}$$

# 単純ベイズ②

## ■ 特徴の独立性を仮定

$$\begin{aligned}\max_{i=1,2,\dots,c} \{p(\omega_i | \mathbf{x})\} &= \max_{i=1,2,\dots,c} \left\{ \frac{p(\mathbf{x} | \omega_i) p(\omega_i)}{p(\mathbf{x})} \right\} \\&= \max_{i=1,2,\dots,c} \{p(\mathbf{x} | \omega_i) p(\omega_i)\} = \max_{i=1,2,\dots,c} \{p(x_1, x_2, \dots, x_d | \omega_i) p(\omega_i)\} \\&= \max_{i=1,2,\dots,c} \{p(x_1 | \omega_i) \times p(x_1 | \omega_i) \cdots \times p(x_d | \omega_i) \times p(\omega_i)\} \\&= \max_{i=1,2,\dots,c} \left\{ \prod_{j=1}^d p(x_j | \omega_i) \times p(\omega_i) \right\}\end{aligned}$$

# 独立性を仮定した条件付き確率の求め方

## ■ 独立性を仮定すると...

- $p(\text{天気=晴れ, 気温=適温, 湿度=適当, 講義=yes} | \text{頭痛=yes})$



- $p(\text{天気=晴れ} | \text{頭痛=yes}) \times p(\text{気温=適温} | \text{頭痛=yes}) \times p(\text{湿度=適当} | \text{頭痛=yes}) \times p(\text{講義=yes} | \text{頭痛=yes})$

- 現実的にはこの方法でも、いずれかの特徴の条件付き確率が0となり計算できない場合も多々ある

# 「天気=晴れ, 気温=適温, 湿度=適当, 講義=yes」において頭痛は？

	天気	気温	湿度	講義	頭痛
1	晴れ	寒い	低い	yes	yes
2	曇り	暑い	高い	no	yes
3	晴れ	暑い	低い	no	no
4	雨	適温	高い	yes	no
5	雨	寒い	高い	no	no
6	晴れ	適温	適当	no	yes
7	曇り	暑い	低い	yes	no
8	雨	寒い	高い	no	no
9	曇り	適温	低い	yes	yes
10	曇り	適温	適当	no	no
11	晴れ	暑い	適当	yes	yes
12	晴れ	適温	高い	no	no
13	雨	暑い	低い	no	no
14	雨	寒い	適当	yes	yes
15	曇り	適温	低い	yes	no
16	晴れ	暑い	適当	no	yes
17	晴れ	寒い	高い	yes	yes
18	曇り	適温	高い	yes	no
19	曇り	適温	適当	no	no
20	雨	寒い	適当	no	yes

## ■ 実習

### ① 事前確率を計算

□  $p(\text{頭痛}=\text{yes}), p(\text{頭痛}=\text{no})$

### ② 条件付き確率を計算

□  $p(\text{天気}=\text{晴れ}|\text{頭痛}=\text{yes})$  など

### ③ ベイズ決定則の適用

# 事前確率の計算

10-25.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K
1		天気	気温	湿度	講義	頭痛					
2	1	晴れ	寒い	低い	yes	yes		事前確率			
3	2	曇り	暑い	高い	no	yes		頭痛=yes			
4	3	晴れ	暑い	低い	no	no		頭痛=no			
5	4	雨	適温	高い	yes	no					
6	5	雨	寒い	高い	no	no					
7	6	晴れ	適温	適当	no	yes		条件付き確率		頭痛	
8	7	曇り	暑い	低い	yes	no			yes	no	
9	8	雨	寒い	高い	no	no		天気	晴れ		
10	9	曇り	適温	低い	yes	yes			曇り		
11	10	曇り	適温	適当	no	no			雨		
12	11	晴れ	暑い	適当	yes	yes		気温	暑い		
13	12	晴れ	適温	高い	no	no			適温		
14	13	雨	暑い	低い	no	no			寒い		
15	14	雨	寒い	適当	yes	yes		湿度	高い		
16	15	曇り	適温	低い	yes	no			適当		
17	16	晴れ	暑い	適当	no	yes			低い		
18	17	晴れ	寒い	高い	yes	yes		講義	yes		
19	18	曇り	適温	高い	yes	no			no		
20	19	曇り	適温	適当	no	no					
21	20	雨	寒い	適当	no	yes		天気=雨, 気温=寒い, 湿度=低い, 講義=yes			
22								事後確率			
23								頭痛=yes			
24								頭痛=no			

事前確率

セルI3  $p(\text{頭痛}=\text{yes})$

セルI4  $p(\text{頭痛}=\text{no})$

# 条件付き確率の計算

## ■ 特徴ごとの条件付き確率を求める

条件付き確率		頭痛	
		yes	no
天気	晴れ		
	曇り		
	雨		
気温	暑い		
	適温		
	寒い		
湿度	高い		
	適当		
	低い		
講義	yes		
	no		

セルJ9

$p(\text{天気=晴れ}|\text{頭痛=yes})$

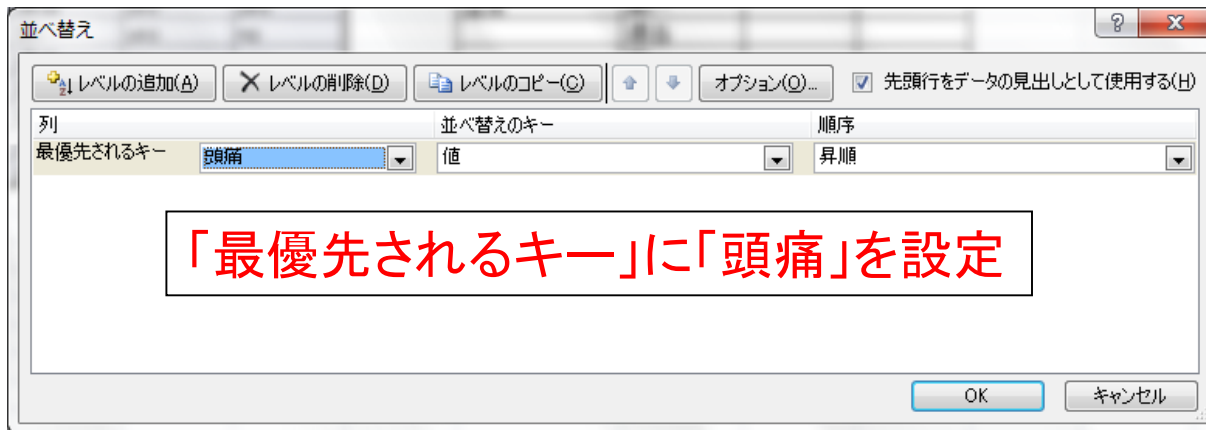
セルK19

$p(\text{講義=no}|\text{頭痛=no})$

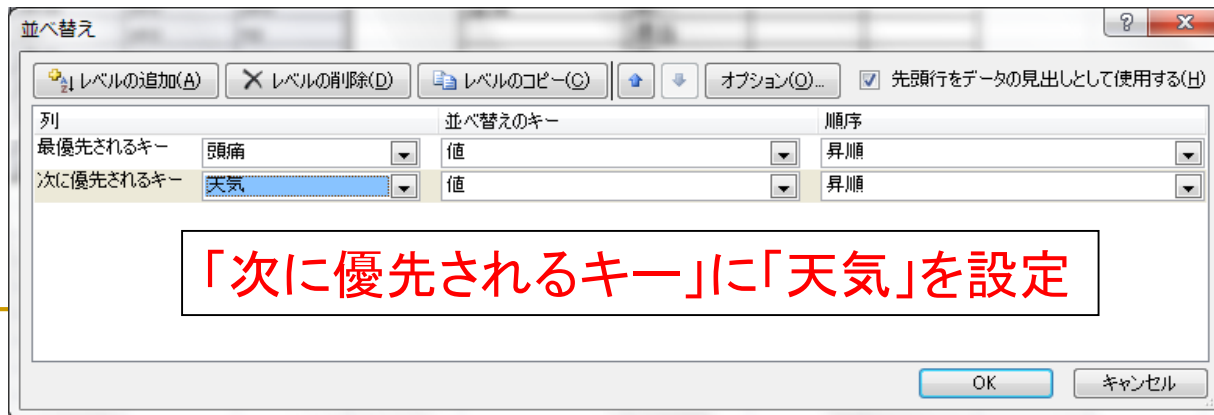
表示形式は「分数」にしている

# 条件付き確率 $p(\text{天気} | \text{頭痛})$ ①

- 「並び変え」を使って数えるとよい
  - 「並び替えとフィルター」→「ユーザー設定の並び替え」



- 「レベルの追加」をクリック



# 条件付き確率 $p(\text{天気} | \text{頭痛})$ ②

10-25.xlsx - Microsoft Excel

	A	B	C	D	E	F
1		天気	気温	湿度	講義	頭痛
2	4	雨	適温	高い	yes	no
3	5	雨	寒い	高い	no	no
4	8	雨	寒い	高い	no	no
5	13	雨	暑い	低い	no	no
6	7	曇り	暑い	低い	yes	no
7	10	曇り	適温	適当	no	no
8	15	曇り	適温	低い	yes	no
9	18	曇り	適温	高い	yes	no
10	19	曇り	適温	適当	no	no
11	3	晴れ	暑い	低い	no	no
12	12	晴れ	適温	高い	no	no
13	14	雨	寒い	適当	yes	yes
14	20	雨	寒い	適当	no	yes
15	2	曇り	暑い	高い	no	yes
16	9	曇り	適温	低い	yes	yes
17	1	晴れ	寒い	低い	yes	yes
18	6	晴れ	適温	適当	no	yes
19	11	晴れ	暑い	適当	yes	yes
20	16	晴れ	暑い	適当	no	yes
21	17	晴れ	寒い	高い	yes	yes

平均: 10.5 データの個数: 120 合計: 210

「頭痛=yes」の場合(9個), 「天気=晴れ」  
となる例(5個)

全ての特徴ごとの条件付き確率を求めなさい



# 条件付き確率 $p(\text{天気} | \text{頭痛})$ ③

	A	B	C	D	E	F
1		天気	気温	湿度	講義	頭痛
2	4	雨	適温	高い	yes	no
3	5	雨	寒い	高い	no	no
4	8	雨	寒い	高い	no	no
5	13	雨	暑い	低い	no	no
6	7	曇り	暑い	低い	yes	no
7	10	曇り	適温	適当	no	no
8	15	曇り	適温	低い	yes	no
9	18	曇り	適温	高い	yes	no
10	19	曇り	適温	適当	no	no
11	3	晴れ	暑い	低い	no	no
12	12	晴れ	適温	高い	no	no
13	14	雨	寒い	適当	yes	yes
14	20	雨	寒い	適当	no	yes
15	2	曇り	暑い	高い	no	yes
16	9	曇り	適温	低い	yes	yes
17	1	晴れ	寒い	低い	yes	yes
18	6	晴れ	適温	適当	no	yes
19	11	晴れ	暑い	適当	yes	yes
20	16	晴れ	暑い	適当	no	yes
21	17	晴れ	寒い	高い	yes	yes

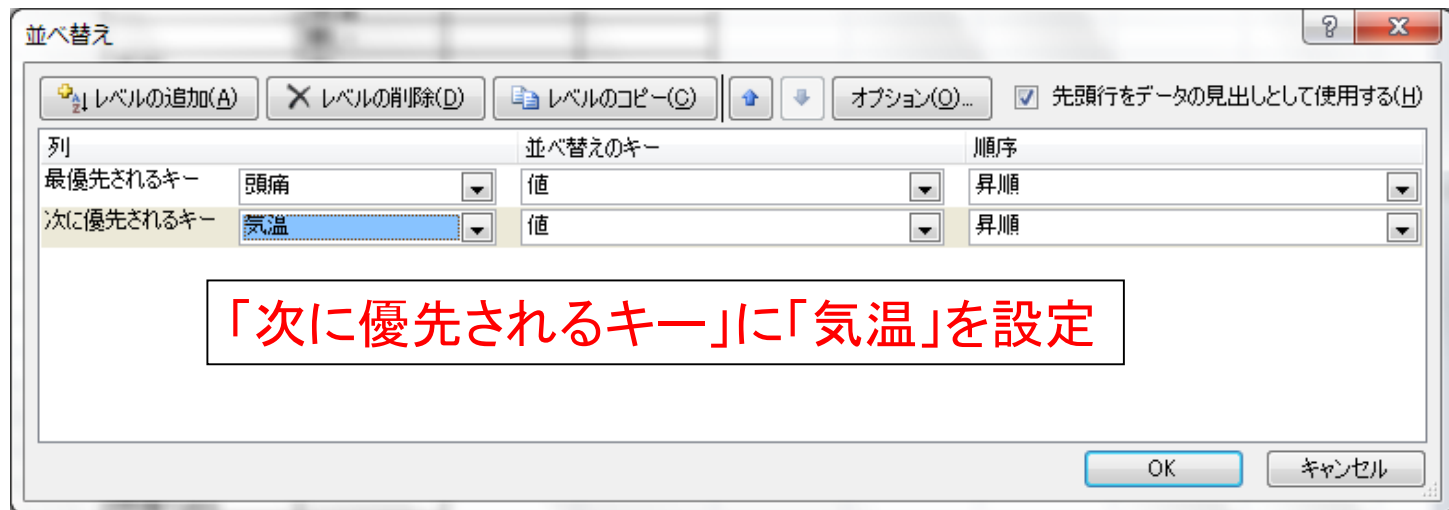
$p(\text{天気=雨} | \text{頭痛=no})$

$p(\text{天気=雨} | \text{頭痛=yes})$

$p(\text{天気=曇り} | \text{頭痛=yes})$

# 条件付き確率 $p(\text{気温} | \text{頭痛})$ ①

- 「並び替えとフィルター」→「ユーザー設定の並び替え」



# 条件付き確率 $p(\text{気温} | \text{頭痛})$ ②

11-1.xlsx - Microsoft Excel

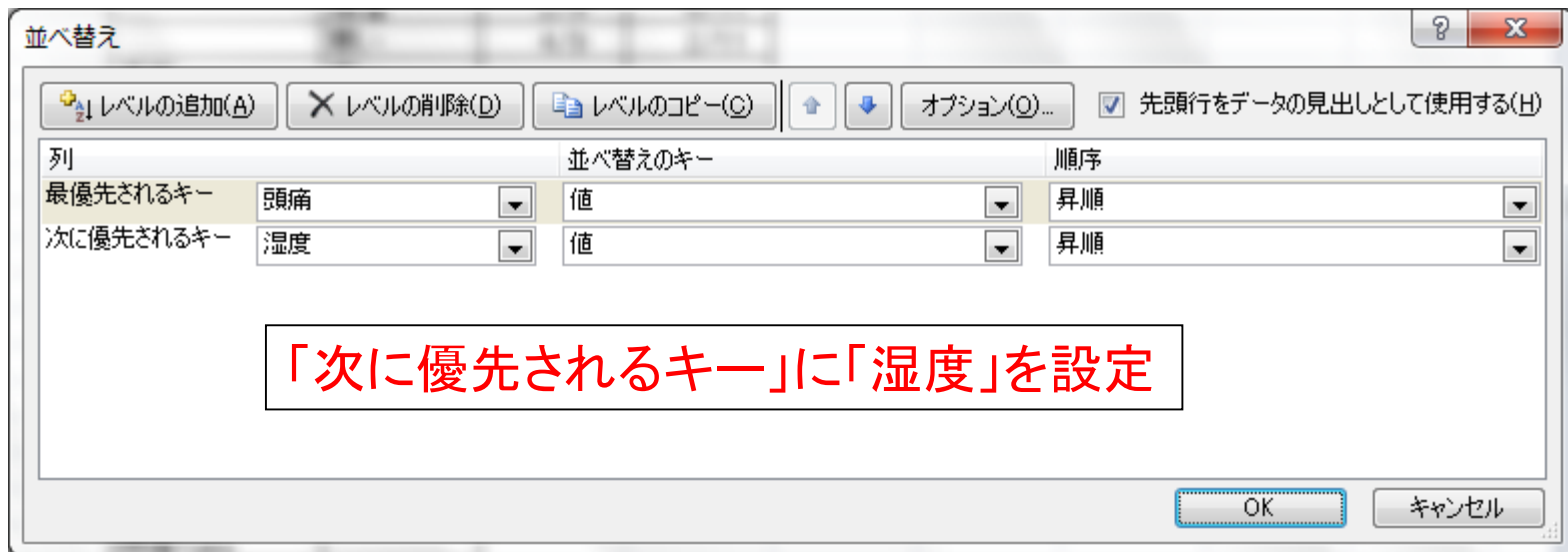
	A	B	C	D	E	F	G
1		天気	気温	湿度	講義	頭痛	
2	1	雨	暑い	低い	no	no	
3	2	曇り	暑い	低い	yes	no	
4	3	晴れ	暑い	低い	no	no	
5	4	雨	寒い	高い	no	no	
6	5	雨	寒い	高い	no	no	
7	6	雨	適温	高い	yes	no	
8	7	曇り	適温	適当	no	no	
9	8	曇り	適温	低い	yes	no	
10	9	曇り	適温	高い	yes	no	
11	10	曇り	適温	適当	no	no	
12	11	晴れ	適温	高い	no	no	
13	12	曇り	暑い	高い	no	yes	
14	13	晴れ	暑い	適当	yes	yes	
15	14	晴れ	暑い	適当	no	yes	
16	15	雨	寒い	適当	yes	yes	
17	16	雨	寒い	適当	no	yes	
18	17	晴れ	寒い	低い	yes	yes	
19	18	晴れ	寒い	高い	yes	yes	
20	19	曇り	適温	低い	yes	yes	
21	20	晴れ	適温	適当	no	yes	
22							

$p(\text{気温}=\text{暑い} | \text{頭痛}=\text{no})$

$p(\text{気温}=\text{適温} | \text{頭痛}=\text{yes})$

# 条件付き確率 $p(\text{湿度} | \text{頭痛})$ ①

- 「並び替えとフィルター」→「ユーザー設定の並び替え」



# 条件付き確率 $p(\text{湿度} | \text{頭痛})$ ②

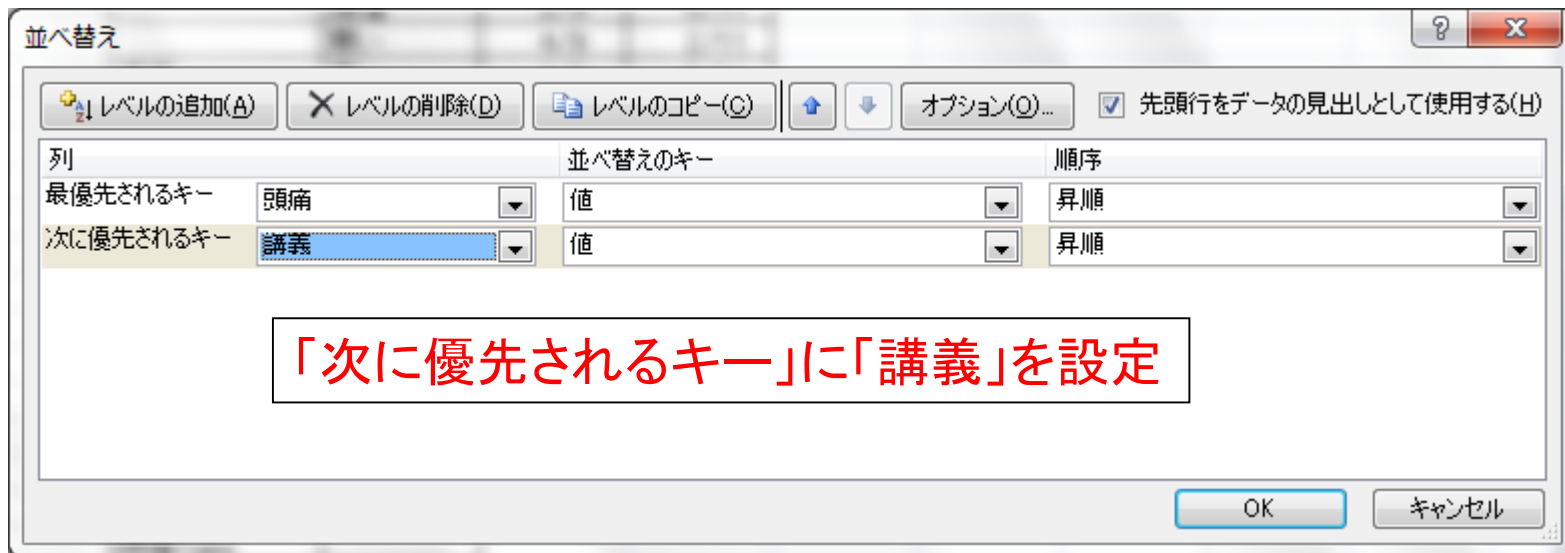
	A	B	C	D	E	F	G
1		天気	気温	湿度	講義	頭痛	
2	1	雨	寒い	高い	no	no	
3	2	雨	寒い	高い	no	no	
4	3	雨	適温	高い	yes	no	
5	4	曇り	適温	高い	yes	no	
6	5	晴れ	適温	高い	no	no	
7	6	曇り	適温	適当	no	no	
8	7	曇り	適温	適当	no	no	
9	8	雨	暑い	低い	no	no	
10	9	曇り	暑い	低い	yes	no	
11	10	晴れ	暑い	低い	no	no	
12	11	曇り	適温	低い	yes	no	
13	12	曇り	暑い	高い	no	yes	
14	13	晴れ	寒い	高い	yes	yes	
15	14	晴れ	暑い	適当	yes	yes	
16	15	晴れ	暑い	適当	no	yes	
17	16	雨	寒い	適当	yes	yes	
18	17	雨	寒い	適当	no	yes	
19	18	晴れ	適温	適当	no	yes	
20	19	晴れ	寒い	低い	yes	yes	
21	20	曇り	適温	低い	yes	yes	
22							

$p(\text{湿度}=\text{高い} | \text{頭痛}=\text{no})$

$p(\text{湿度}=\text{低い} | \text{頭痛}=\text{yes})$

# 条件付き確率 $p(\text{講義} | \text{頭痛})$ ①

- 「並び替えとフィルター」→「ユーザー設定の並び替え」



# 条件付き確率 $p(\text{講義} | \text{頭痛})$ ②

11-1.xlsx - Microsoft Excel

	A	B	C	D	E	F	G
1		天気	気温	湿度	講義	頭痛	
2	1	雨	寒い	高い	no	no	
3	2	雨	寒い	高い	no	no	
4	3	晴れ	適温	高い	no	no	
5	4	曇り	適温	適当	no	no	
6	5	曇り	適温	適当	no	no	
7	6	雨	暑い	低い	no	no	
8	7	晴れ	暑い	低い	no	no	
9	8	雨	適温	高い	yes	no	
10	9	曇り	適温	高い	yes	no	
11	10	曇り	暑い	低い	yes	no	
12	11	曇り	適温	低い	yes	no	
13	12	曇り	暑い	高い	no	yes	
14	13	晴れ	暑い	適当	no	yes	
15	14	雨	寒い	適当	no	yes	
16	15	晴れ	適温	適当	no	yes	
17	16	晴れ	寒い	高い	yes	yes	
18	17	晴れ	暑い	適当	yes	yes	
19	18	雨	寒い	適当	yes	yes	
20	19	晴れ	寒い	低い	yes	yes	
21	20	曇り	適温	低い	yes	yes	
22							

$p(\text{講義}=\text{no} | \text{頭痛}=\text{no})$

$p(\text{講義}=\text{yes} | \text{頭痛}=\text{yes})$

# 事後確率の計算①

セルI22

$p(\text{頭痛}=\text{yes}|\text{天気}=\text{晴れ}, \text{気温}=\text{適温}, \text{湿度}=\text{適当}, \text{講義}=\text{yes})$

天気=晴れ, 気温=適温, 湿度=適当, 講義=yes			
事後確率			
頭痛=yes			
頭痛=no			

ベイズの定理



■  $p(\text{天気}=\text{晴れ}, \text{気温}=\text{適温}, \text{湿度}=\text{適当}, \text{講義}=\text{yes}|\text{頭痛}=\text{yes}) \times p(\text{頭痛}=\text{yes})$



独立性の仮定

$p(\text{天気}=\text{晴れ}|\text{頭痛}=\text{yes}) \times p(\text{気温}=\text{適温}|\text{頭痛}=\text{yes}) \times$

J9

J13

$p(\text{湿度}=\text{適当}|\text{頭痛}=\text{yes}) \times p(\text{講義}=\text{yes}|\text{頭痛}=\text{yes}) \times p(\text{頭痛}=\text{yes})$

J16

J18

I3



# 事後確率の計算②

セルI23

$p(\text{頭痛=no} | \text{天気=晴れ, 気温=適温, 湿度=適当, 講義=yes})$

天気=晴れ, 気温=適温, 湿度=適当, 講義=yes			
事後確率			
頭痛=yes			
頭痛=no			

ベイズの定理



■  $p(\text{天気=晴れ, 気温=適温, 湿度=適当, 講義=yes} | \text{頭痛=no}) \times p(\text{頭痛=no})$



独立性の仮定

$p(\text{天気=晴れ} | \text{頭痛=no}) \times p(\text{気温=適温} | \text{頭痛=no}) \times$

K9

K13

$p(\text{湿度=適当} | \text{頭痛=no}) \times p(\text{講義=yes} | \text{頭痛=no}) \times p(\text{頭痛=no})$

K16

K18

I4

# ベイズ決定則①

- 事後確率の大小によって判定
  - $p(\text{頭痛}=\text{yes}|\text{天気}=\text{晴れ}, \text{気温}=\text{適温}, \text{湿度}=\text{適当}, \text{講義}=\text{yes})$
  - $p(\text{頭痛}=\text{no}|\text{天気}=\text{晴れ}, \text{気温}=\text{適温}, \text{湿度}=\text{適当}, \text{講義}=\text{yes})$
- 「天気=雨, 気温=寒い, 湿度=低い, 講義=yes」の場合も判定しなさい

# 事後確率の計算③

セルI28

$p(\text{頭痛}=\text{yes}|\text{天気}=\text{雨}, \text{気温}=\text{寒い}, \text{湿度}=\text{低い}, \text{講義}=\text{yes})$

ベイズの定理



■  $p(\text{天気}=\text{雨}, \text{気温}=\text{寒い}, \text{湿度}=\text{低い}, \text{講義}=\text{yes}|\text{頭痛}=\text{yes}) \times p(\text{頭痛}=\text{yes})$



独立性の仮定

$p(\text{天気}=\text{雨}|\text{頭痛}=\text{yes}) \times p(\text{気温}=\text{寒い}|\text{頭痛}=\text{yes}) \times$

J11

J14

$p(\text{湿度}=\text{低い}|\text{頭痛}=\text{yes}) \times p(\text{講義}=\text{yes}|\text{頭痛}=\text{yes}) \times p(\text{頭痛}=\text{yes})$

J17

J18

I3

# 事後確率の計算④

セルI29

$p(\text{頭痛=no} | \text{天気=雨, 気温=寒い, 湿度=低い, 講義=yes})$

ベイズの定理



■  $p(\text{天気=雨, 気温=寒い, 湿度=低い, 講義=yes} | \text{頭痛=no}) \times p(\text{頭痛=no})$



独立性の仮定

$p(\text{天気=雨} | \text{頭痛=no}) \times p(\text{気温=寒い} | \text{頭痛=no}) \times$

K11

K14

$p(\text{湿度=低い} | \text{頭痛=no}) \times p(\text{講義=yes} | \text{頭痛=no}) \times p(\text{頭痛=no})$

K17

K18

I4

# ベイズ決定則②

- 最も頭痛が起きる場合の特徴( $a_1, a_2, a_3, a_4$ )を求めるためにはどうすればよいか. 考えなさい.
- $p(\text{頭痛}=\text{yes}|\text{天気}=a_1, \text{気温}=a_2, \text{湿度}=a_3, \text{講義}=a_4)$

# 単純ベイズ決定則のプログラム

# 単純ベイズ決定則 (Cancer\_NB.py)

## ■ 乳がんの分類問題

### □ Breast cancer dataset

用途	クラス分類
データ数	569
特徴量	30
目的変数	2

クラス名	データ数
malignant	212
benign	357

### □ ITCのPCでは, Cancer\_NB-ITC.py を実行して下さい

# Cancer\_NB.pyで用いる確率密度関数

## ■ GaussianNBでの正規分布

- 平均ベクトル:  $\mathbf{m}_i$
- 分散行列:  $\Sigma_i$

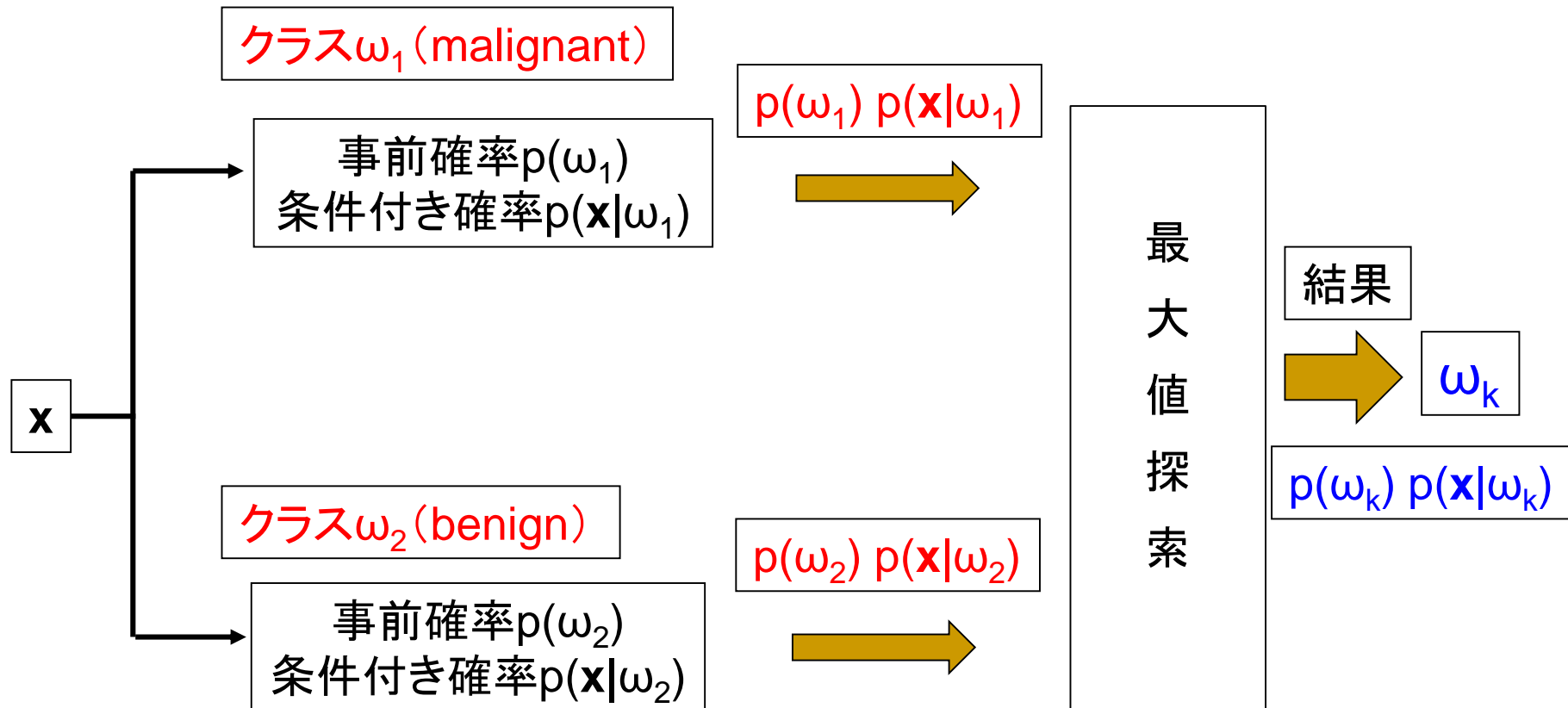
$$\Sigma_i = \begin{pmatrix} \sigma_{i1} & 0 & \cdots & 0 \\ 0 & \sigma_{i2} & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & \sigma_{id} \end{pmatrix}$$

分散行列(共分散は0)

$$p(\mathbf{x} | \omega_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i)\right)$$



# 単純ベイズ決定則



$$p(\mathbf{x} | \omega_i) = -\frac{1}{2} (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i) - \frac{1}{2} \log |\Sigma_i| + \log p(\omega_i)$$

# Cancer\_NB.py

```
import numpy
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import classification_report, accuracy_score,
confusion_matrix
```

単純ベイズのためにimportが必要

# データのロード

breast cancerデータセットの読み込み

```
cancer = datasets.load_breast_cancer()
```

# 種類( malignant, benign )

```
name = cancer.target_names
```

```
label = cancer.target
```

label: 正解ラベル  
malignant → 0  
benign → 1

個数  
569

# 特徴量

```
feature_names = cancer.feature_names
```

```
data = cancer.data
```

data  
特徴量

大きさ  
(569,30)

# データ数, 特徴量の個数

```
total, feature_count = data.shape
```

total  
データ数

feature\_count  
特徴量の個数

# 学習データ, テストデータ

ホールドアウト法

```
train_data, test_data, train_label, test_label = train_test_split(data, label,  
test_size=0.5, random_state=None)
```

priors: 事前確率  
Noneの場合, 初期設定しない (default)

```
model = GaussianNB(priors=None, var_smoothing=1e-9)
```

var\_smoothing: 分散の最小値  
分散が0となるのを防ぐ (defaultは1e-9)

# 学習

```
model.fit(train_data, train_label)
```

# 予測

```
predict = model.predict(test_data)
```

```
predict_proba = model.predict_proba(test_data)
```

```
from sklearn.naive_bayes import GaussianNB  
GaussianNB(priors=None, var_smoothing=1e-9)
```

priors: 事前確率  
Noneの場合, 事前確率を求める(default)

var\_smoothing: 分散の最小値  
分散が0となるのを防ぐ(defaultは1e-9)

```
model = GaussianNB()
```

```
model = GaussianNB(priors=None, var_smoothing=1e-9)
```



```
model = GaussianNB(priors=[0.5,0.5], var_smoothing=1e-9)
```

事前確率を二クラスとも同一にする  
(ベイズ決定則ではなく最尤法)

```
print( " malignant benign -> 予測結果 : 正解ラベル" )  
for i in range(20):  
    print( " {0:5.3f} {1:5.3f} -> {2:2d} : {3:2d}".format  
    ( predict_proba[i][0] , predict_proba[i][1] , predict[i] , test_label[i] ) )
```

クラス1 (malignant)  
の事後確率

クラス2 (benign)  
の事後確率

予測結果

正解ラベル

```
predict_proba[i][0] > predict_proba[i][1]  
→ i番目のデータはクラス1 (malignant)と予測
```

```
predict_proba[i][0] < predict_proba[i][1]  
→ i番目のデータはクラス2 (benign)と予測
```

## # 事前確率, 平均値, 分散の表示

```
for i in range(len(name)):
```

```
    print( "¥n [ {} ]".format( name[i] ) )
```

```
    print( " 事前確率 : {0:.3f}".format( model.class_prior_[i] ) )
```

```
    print( " 平均値  :" )
```

```
    for j in range(feature_count):
```

```
        if j % 10 == 0 and j != 0 : print()
```

```
        print( " {0:8.3f}".format( model.theta_[i][j] ) , end="" )
```

```
    print( "¥n 分散  :" )
```

```
    for j in range(feature_count):
```

```
        if j % 10 == 0 and j != 0 : print()
```

```
        print( " {0:8.3f} ".format( model.sigma_[i][j] ) , end="" )
```

```
    print()
```

`class_prior_`: 事前確率  
個数: クラス数

`theta_`: 平均ベクトル  
(クラス数, 次元数)

`sigma_`: 分散ベクトル  
(クラス数, 次元数)

## 推定した正規分布のパラメータ

`theta_`: 平均ベクトル

`sigma_`: 分散ベクトル

```
print( "¥n [ 予測結果 ]" )
```

```
print( classification_report(test_label, predict) )
```

accuracy  
precision  
recall  
F値

```
print( "¥n [ 正解率 ]" )
```

```
print( accuracy_score(test_label, predict) )
```

accuracyの表示

```
print( "¥n [ 混同行列 ]" )
```

```
print( confusion_matrix(test_label, predict) )
```

混同行列の表示

# 実行結果①

クラス1 (malignant)  
の事後確率

クラス2 (benign)  
の事後確率

```
C:\Windows\system32\cmd.exe
malignant benign -> 予測結果 : 正解ラベル
0.000 1.000 -> 1 : 0
0.009 0.991 -> 1 : 1
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
0.261 0.739 -> 1 : 0
0.148 0.852 -> 1 : 1
0.000 1.000 -> 1 : 1
1.000 0.000 -> 0 : 0
1.000 0.000 -> 0 : 0
1.000 0.000 -> 0 : 0
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
1.000 0.000 -> 0 : 0
0.000 1.000 -> 1 : 1
1.000 0.000 -> 0 : 0
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
0.000 1.000 -> 1 : 1
```

クラス1の事後確率 > クラス2の事後確率  
→ クラス1 (malignant)

クラス1の事後確率 < クラス2の事後確率  
→ クラス2 (benign)

正解ラベル

予測結果



## 実行結果②

事前確率

平均ベクトル(30次元)

分散ベクトル(30次元)

[ malignant ]										
事前確率	: 0.370									
平均値	:									
17.788	21.980	117.781	1015.133	0.103	0.149	0.168	0.091	0.194	0.062	
0.630	1.215	4.505	76.677	0.007	0.033	0.044	0.015	0.019	0.004	
21.473	29.504	144.350	1468.202	0.143	0.378	0.458	0.187	0.318	0.090	
分散	:									
10.503	13.666	493.060	141971.782	0.000	0.003	0.007	0.002	0.001	0.000	
0.137	0.218	7.032	4351.643	0.000	0.001	0.001	0.000	0.000	0.000	
19.978	23.967	967.705	404542.431	0.001	0.026	0.031	0.003	0.005	0.001	
[ benign ]										
事前確率	: 0.630									
平均値	:									
12.235	17.657	78.656	468.958	0.092	0.081	0.048	0.027	0.173	0.063	
0.278	1.176	1.965	20.948	0.007	0.022	0.027	0.010	0.020	0.004	
13.473	23.103	87.557	566.516	0.124	0.187	0.175	0.076	0.269	0.080	
分散	:									
3.058	17.463	133.885	17745.179	0.001	0.002	0.002	0.001	0.001	0.000	
0.011	0.345	0.556	66.439	0.000	0.001	0.001	0.000	0.000	0.000	
3.839	31.800	176.110	26691.245	0.001	0.009	0.023	0.002	0.002	0.001	

# 実行結果③

```
C:\Windows\system32\cmd.exe

[ 予測結果 ]
      precision    recall  f1-score   support

     0       0.94      0.87      0.90       107
     1       0.92      0.97      0.95       178

 accuracy          0.93       285
 macro avg       0.93      0.92      0.92       285
weighted avg       0.93      0.93      0.93       285

[ 正解率 ]
0.9298245614035088

[ 混同行列 ]
[[ 93  14]
 [  6 172]]
```

accuracy  
precision  
recall  
F値

accuracyの表示

混同行列の表示

# 参考文献①

- 舟久保登：パターン認識，共立出版（1991）
- 石井健一郎他：わかりやすいパターン認識，オーム社（1998）
- 杉山将：統計的機械学習，オーム社（2009）
- 浜本義彦：統計的パターン認識入門，森北出版（2009）
- Richard O. Duda他：パターン識別，アドコム・メディア（2009）

## 参考文献②

- GaussianNB
- [https://scikit-learn.org/stable/modules/generated/sklearn.naive\\_bayes.GaussianNB.html](https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html)