はじめてのパターン認識
第3章･ベイズの識別規則

10T4065Y
山木翔馬

最大事後確率基準

ベイズの定理

クラス条件付き確率
（尤度）

$$P\left(C_i \middle| x\right) = \frac{p\left(x \middle| C_i\right)}{p\left(x\right)} \times P\left(C_i\right)$$

事後確率　　周辺確率　　事前確率

これが最大となるクラスに分類する

観測データ x が与えられた下で、それが
クラス Ci に属する条件付き確率

事後確率

$$\boxed{P\left(C_i \middle| x\right)} = \frac{p\left(x \middle| C_i\right)}{p\left(x\right)} \times \boxed{P\left(C_i\right)}$$

事前確率

クラス Ci の生起確率で、
観測する前から分かっている確率

クラスが与えられた下での観測データ x の確率分布

クラス条件付き確率（尤度）

$$P\left(C_i \middle| x\right) = \frac{\boxed{p\left(x \middle| C_i\right)}}{\boxed{p\left(x\right)}} \times P\left(C_i\right)$$

周辺確率

観測データ x の生起確率で、

$$p\left(x\right) = \sum_{i=1}^{K} p\left(C_i, x\right)$$

で得られる ➡ 周辺化

ベイズの識別規則

クラス Ci と Cj の識別境界

$$P(C_i|x)=\frac{p(x|C_i)P(C_i)}{p(x)}=\frac{p(x|C_j)P(C_j)}{p(x)}=P(C_j|x)$$

共通しているので・・・

$$\boxed{\text{識別クラス}=\arg\max_i p(x|C_i)P(C_i)}$$

ベイズの識別規則の例

	サンプル数	喫煙する人（S = 1）	飲酒をする人（T = 1）
健康な人（G = 1）	800人	320人	640人
健康でない人（G = 0）	200人	160人	40人

特徴 (s, t) のすべての組み合わせについて

$$P(G|S,T) = \frac{P(S,T|G)\,P(G)}{P(S,T)}$$

を求める

各クラスの事前確率

$$P(G=1)=\frac{800}{1000}=\frac{4}{5}, P(G=0)=\frac{200}{1000}=\frac{1}{5}$$

クラス条件付き確率

（SとTの間に条件付き独立が成立すると仮定）

$$P(S,T|G)=P(S|G)P(T|G)$$

（例）喫煙する人の
クラス条件付き確率

$$P(S=1|G=1)=\frac{320}{800}=\frac{2}{5}$$

周辺確率

$$P(S,T)=P(S,T,G=1)+P(S,T,G=0)$$

クラス条件付き確率、同時確率、周辺確率の計算結果

	(S, T)			
	(1, 1)	(0, 1)	(1, 0)	(0, 0)
P(S, T | G = 1)	8 / 25	12 / 25	2 / 25	3 / 25
P(S, T | G = 0)	4 / 25	1 / 25	16 / 25	4 / 25
P(S, T, G = 1)	32 / 125	48 / 125	8 / 125	12 / 125
P(S, T, G = 0)	4 / 125	1 / 125	16 / 125	4 / 125
P(S, T)	36 / 125	49 / 125	24 / 125	16 / 125

※ $$P\left(S=1,T=1,G=1\right)=P\left(S=1,T=1\middle|G=1\right)\cdot P\left(G=1\right)$$

$$=\frac{8}{25}\cdot\frac{4}{5}=\frac{32}{125}$$

事後確率の計算結果と健康か否かの判断

	(S, T)			
	(1, 1)	(0, 1)	(1, 0)	(0, 0)
P(G = 1 | S, T)	8 / 9	48 / 49	1 / 3	3 / 4
P(G = 0 | S, T)	1 / 9	1 / 49	2 / 3	1 / 4
判断	G = 1	G = 1	G = 0	G = 1

尤度比によるクラス識別

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{\cancel{p(x)}} = \frac{p(x|C_j)P(C_j)}{\cancel{p(x)}} = P(C_j|x)$$

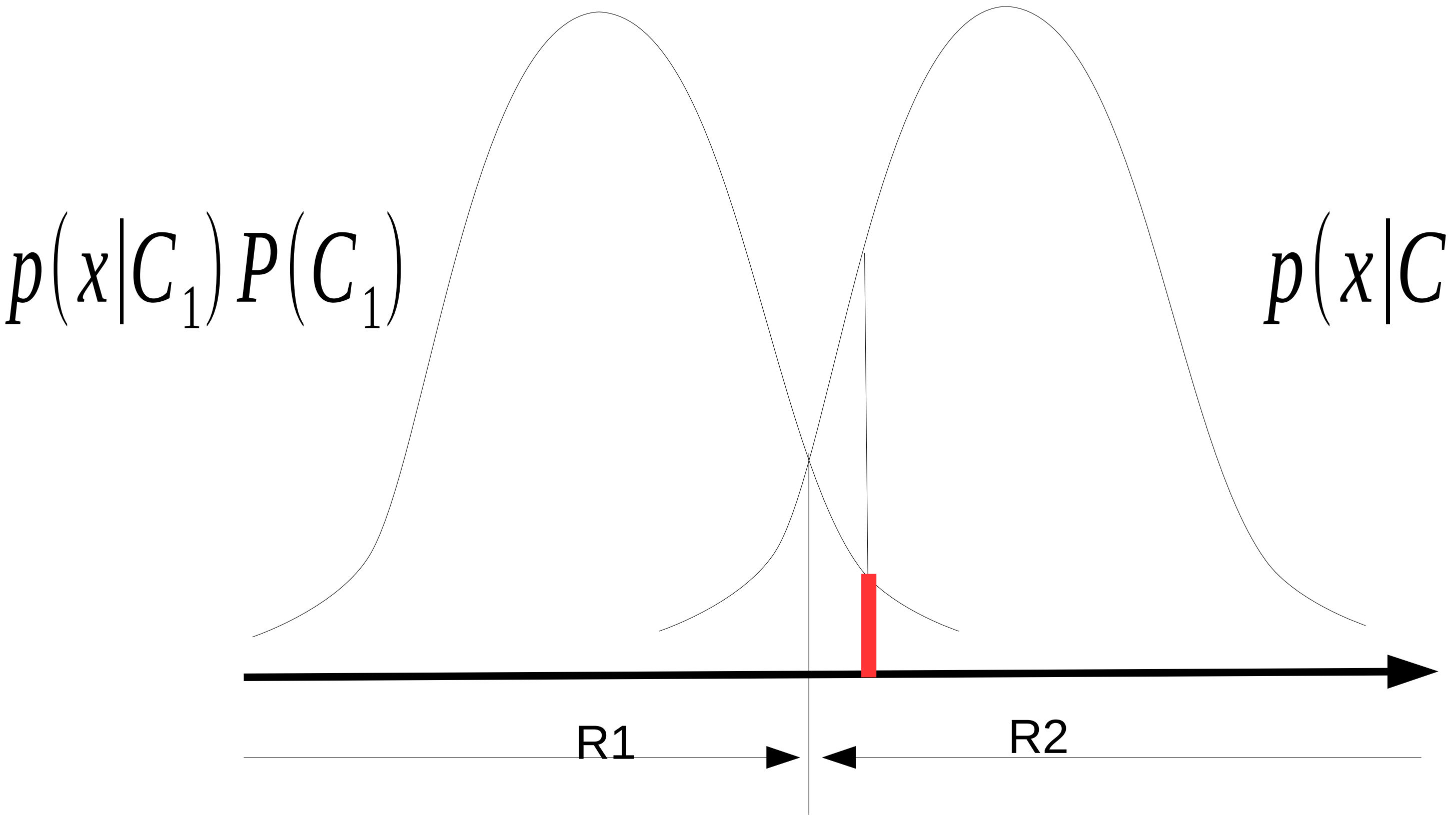
$$\downarrow$$

$$p(x|C_i)P(C_i) \left\{ \begin{matrix} > \\ < \end{matrix} \right\} p(x|C_j)P(C_j) \left\{ \begin{matrix} => C_i \\ => C_j \end{matrix} \right\}$$

$$\downarrow$$

尤度比 $$\frac{p(x|C_i)}{p(x|C_j)} \left\{ \begin{matrix} > \\ < \end{matrix} \right\} \frac{P(C_i)}{P(C_j)} = h_{ij} \left\{ \begin{matrix} => C_i \\ => C_j \end{matrix} \right\}$$

条件付きベイズ誤り率、ベイズ誤り率

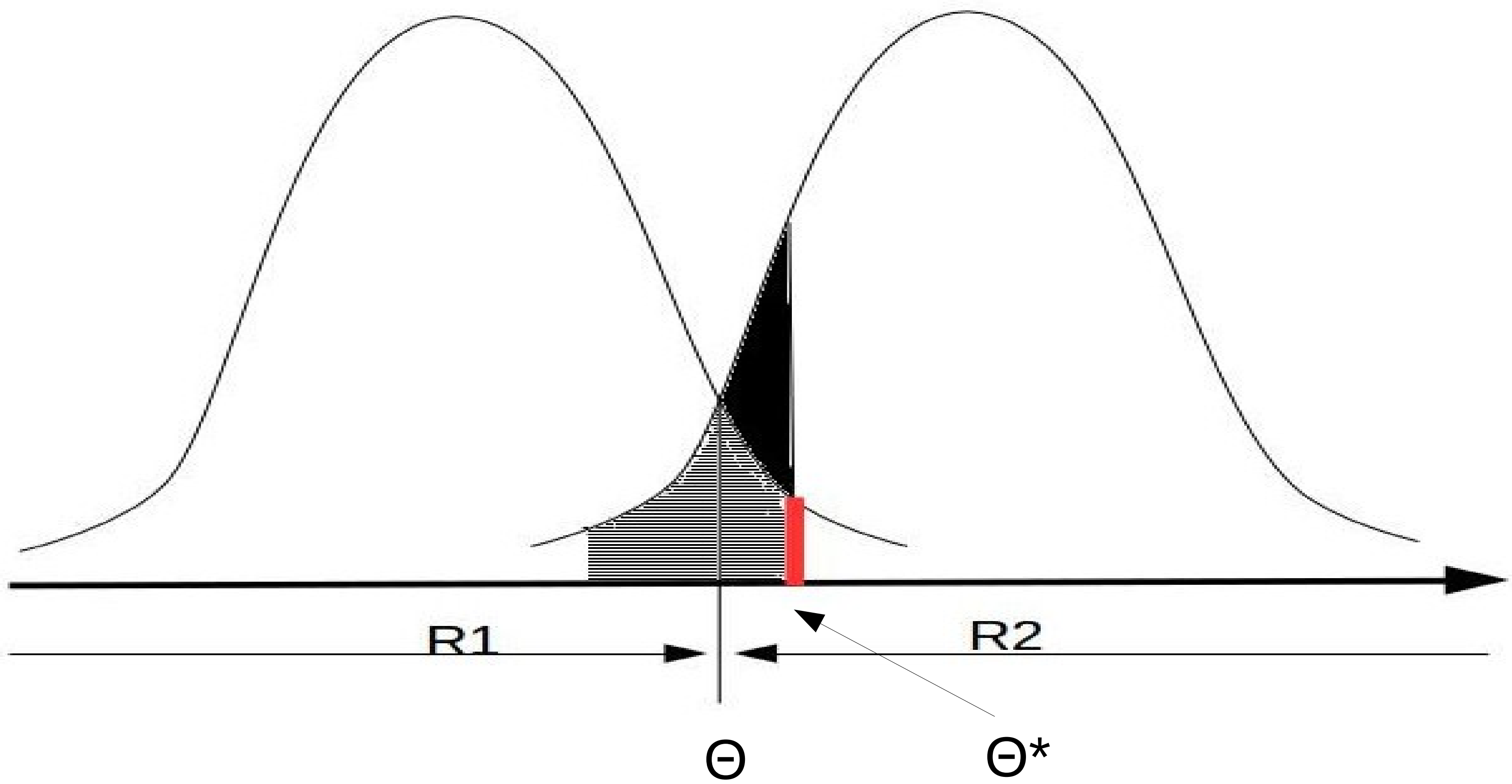


条件付きベイズ誤り率　$\varepsilon(x) = min[P(C_1|x), P(C_2|x)]$

ベイズ誤り率　$\varepsilon^* = E\{\varepsilon(x)\} = \int_{R_1+R_2} \varepsilon(x) p(x) dx$

$$= \int_{R_2} p(x|C_1) P(C_1) dx + \int_{R_1} p(x|C_2) P(C_2) dx$$

ベイズの識別規則は誤り率最小



識別境界がΘ*にずれると、誤り率が増加
　→ベイズの識別規則は誤り率が最小になる

最小損失基準に基づくベイズの識別規則

健康な人が病気と判断される時の危険性 ＜ 病気の人が健康と判断される時の危険性

危険性を加味した識別規則

L_{ij}　真のクラスがCjのとき、Ciと判断することによって被る損失

Lijを要素とするK×Kの行列を損失行列

観測データxをクラスCiと判断した時に被る損失は

$$r\left(C_i|x\right)=\sum_{k=1}^{K} L_{ik} P\left(C_k|x\right)$$

つまり、識別規則は
損失の最も小さなクラスに識別すること

$$\boxed{\text{識別クラス}=\underset{i}{\arg\min}\, r\left(C_i|x\right)}$$

損失の期待値

入力xが与えられたときに被る損失は

$$r(x) = min[r(C_1|x), r(C_2|x)]$$

クラス1、クラス2に識別される領域全体に渡る損失の期待値は

$$r = E\{r(x)\} = \int_{R_1+R_2} min[r(C_1|x), r(C_2|x)]p(x)dx$$

$$= \int_{R_1} (L_{11} p(x|C_1)P(C_1) + L_{12} p(x|C_2)P(C_2))dx$$

$$+ \int_{R_2} (L_{21} p(x|C_1)P(C_1) + L_{22} p(x|C_2)P(C_2))dx$$

識別規則

一般に$L_{ii} \leq L_{ij} \left(i \neq j \right)$

$$\left(L_{21} - L_{11} \right) p\left(x \middle| C_1 \right) P\left(C_1 \right) \begin{Bmatrix} > \\ < \end{Bmatrix} \left(L_{12} - L_{22} \right) p\left(x \middle| C_2 \right) P\left(C_2 \right) \begin{Bmatrix} => C_1 \\ => C_2 \end{Bmatrix}$$

尤度比を用いた場合

$$\frac{p\left(x \middle| C_1 \right)}{p\left(x \middle| C_2 \right)} \begin{Bmatrix} > \\ < \end{Bmatrix} \frac{\left(L_{12} - L_{22} \right) P\left(C_2 \right)}{\left(L_{21} - L_{11} \right) P\left(C_1 \right)} \begin{Bmatrix} => C_1 \\ => C_2 \end{Bmatrix}$$

リジェクト

・誤り率が大きな領域では判断を避けること

$$\frac{p(x|C_1)}{p(x|C_2)} \leq \frac{1-t}{t}\frac{P(C_2)}{P(C_1)}$$

・上を満たす領域がクラス1個のリジェクト領域

・しきい値 t を下げればリジェクト率は増加し、
誤って認識する確率も減少する

ROC曲線

- 受信者動作特性曲線
- 事前確率や尤度、識別境界のような情報を必要としない性能評価法