はじめてのパターン認識
第３章　ベイズの識別規則

茨城大学大学院理工学研究科

菊池裕紀

3.1 ベイズの識別規則(1)

- 最大事後確率基準

観測データ：x

識別クラス：$C_i \ (i = 1,2,\ldots,K)$

$$\underbrace{P(C_i|x)}_{\text{事後確率}} = \frac{\overbrace{p(\boldsymbol{x}|C_i)}^{\text{尤度}}}{\underbrace{p(\boldsymbol{x})}_{\text{周辺確率}}} \times \underbrace{P(C_i)}_{\text{事前確率}}$$

・・・ベイズの定理

事後確率が最大となるクラスに観測データを分類

3.1 ベイズの識別規則(2)

- クラスの識別境界

  事後確率が等しくなるところ

$$P(C_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_i)P(C_i)}{\underbrace{p(\boldsymbol{x})}_{\text{共通}}} = \frac{p(\boldsymbol{x}|C_j)P(C_j)}{\underbrace{p(\boldsymbol{x})}_{\text{共通}}} = P(C_j|\boldsymbol{x})$$

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

3.1 ベイズの識別規則(3)

- ベイズの識別規則の例

	サンプル数	喫煙する人 (S = 1)	飲酒する人 (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

目的：この町の他の人の喫煙と飲酒の有無を情報から、その人の健康状態を予測する

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

3.1 ベイズの識別規則(4)

- 事前確率

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

- クラスの条件付き確率（S,Tの条件付き独立を仮定

$$P(S=1|G=1)=\frac{320}{800}=\frac{2}{5},\quad \mathrm{P}(S=0|G=1)=\frac{480}{800}=\frac{3}{5}$$

$$P(S=1|G=0)=\frac{160}{200}=\frac{4}{5},\qquad \mathrm{P}(S=0|G=0)=\frac{40}{200}=\frac{1}{5}$$

- 周辺確率

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

3.1 ベイズの識別規則(4)

	(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

クラス付き条件確率と事前確率の積の比較

$$P(\boldsymbol{x}|C_i)P(C_i)\{>,<\}p(\boldsymbol{x}|C_j)P(C_j) \rightarrow C_i \ or \ C_j$$

尤度比

$$\frac{P(\boldsymbol{x}|C_i)}{P(\boldsymbol{x}|C_j)}\{>,<\}\frac{P(C_j)}{P(C_i)} = h_{ij} \rightarrow C_i \ or \ C_j$$

3.2 受信者動作特性曲線

- 受信者動作特性曲線（ROC曲線）

性能評価法として利用されているもの

以下、2クラス問題の場合...

クラスに属していると判断　　：p(positive)

クラスに属していないと判断　：n(negative)

$\{p^*, n^*\}$で $\boldsymbol{x}$ の真のクラスを表すこととする。

3.2 受信者動作特性曲線(2)

		識別クラス		行和
		P	n	
真のクラス	p^*	True Positive 真陽性	False Negative 偽陰性	P=TP+FN
	n^*	False Positive 偽陽性	True Negative 真陰性	N=FP+TN