

# 「パターン認識と機械学習」

## 1-5 決定理論

### 1-5-1 誤識別率の最小化

新納浩幸

# 決定理論

訓練データから  $p(\mathbf{x}, \mathbf{t})$  を決めるのが**推論**



現実のタスクでは **これ** を使って、ある行動を行う  
その際の意思決定の理論が**決定理論**

# 例


## 医療診断問題

$\mathbf{x}$  入力、画像ピクセル強度の集合

$C_1$  : ガンのクラス       $C_2$  : 非ガンのクラス

出力

$t = 0$        $\mathbf{x} \in C_1$

$t = 1$        $\mathbf{x} \in C_2$

    $p(\mathbf{x}, t)$    を決めることが大事

# 直感的な解

$$p(C_1 | \mathbf{x}) \quad \text{or} \quad p(C_2 | \mathbf{x})$$

大きい方を選べば良い...

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

右辺の分子の大きい方を選べば良い...

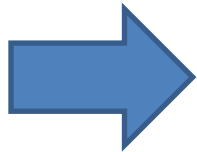
# 誤識別率

$R_k$  :  $\mathbf{x}$  を  $C_k$  と識別する領域 **決定領域**

**誤識別率**

$$p(\text{Error}) = \int_{R_1} p(\mathbf{x}, C_2) d\mathbf{x} + \int_{R_2} p(\mathbf{x}, C_1) d\mathbf{x}$$

これを最小にするようにクラスを割り振れば良い



先の直感的な解が正しい