パターン認識と機械学習
1.5.4 推論と決定

吉田拓夢

クラス分類問題

推論段階(inference stage)

・訓練データ→モデル学習：$p(C_k|x)$

決定段階(decision stage)

・事後確率を使い最適なクラスに割当

あるいは………

両方の問題を同時に解いて入力xから直接

決定関数を単に学習

識別関数(discriminant function)

アプローチ(a) 生成モデル(generative model)

- クラスC_kごとに条件付き密度$p(x|C_k)$を決める
  事前クラス確率$p(C_k)$もそれぞれ求める
- クラス事後確率$p(C_k|x)$を求める

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

$p(C_k|x)$は分子に現れる量で計算できる

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

アプローチ(a) 生成モデル(generative model)

前述に等価な手法

- 同時分布p(x,Ck)を直接モデル化→規格化

決定理論によって……

事後確率が分かる

→新たな入力xのクラス属性を決められる

> 生成モデル
>
> 出力だけでなく入力の分布もモデル化
>
> モデルからのサンプリングで入力空間で人工データ点を生成可能

アプローチ(b) 識別モデル(discriminative model)

- クラス事後確率$p(C_k|x)$を求める

- 決定理論でそれぞれの新たなxをクラスの１つに割当

識別(判別)モデル

事後確率を直接モデル化

<u>アプローチ (c)</u>

- <u>識別関数</u>を見つける

　　→ 各入力xから直接クラスラベルに写像する関数$f(x)$

例えば……

2クラス問題では、$f(\cdot)$は2値で、$f=0$はC_1を表し、

$f=1$はクラスC_2を表す(この場合、確率は出てこない)

アプローチ(a)の得失

- 3つの中で最も手間がかかる←xとC_kの両方の同時分布

x は高次元→十分な精度の$p(x|C_k)$には多くの訓練集合が必要
$p(C_k)$は多くの場合、単に各クラスの訓練集合の比率で推定できる

- データの周辺分布p(x)が決められる

予測の精度が低いと考えられる

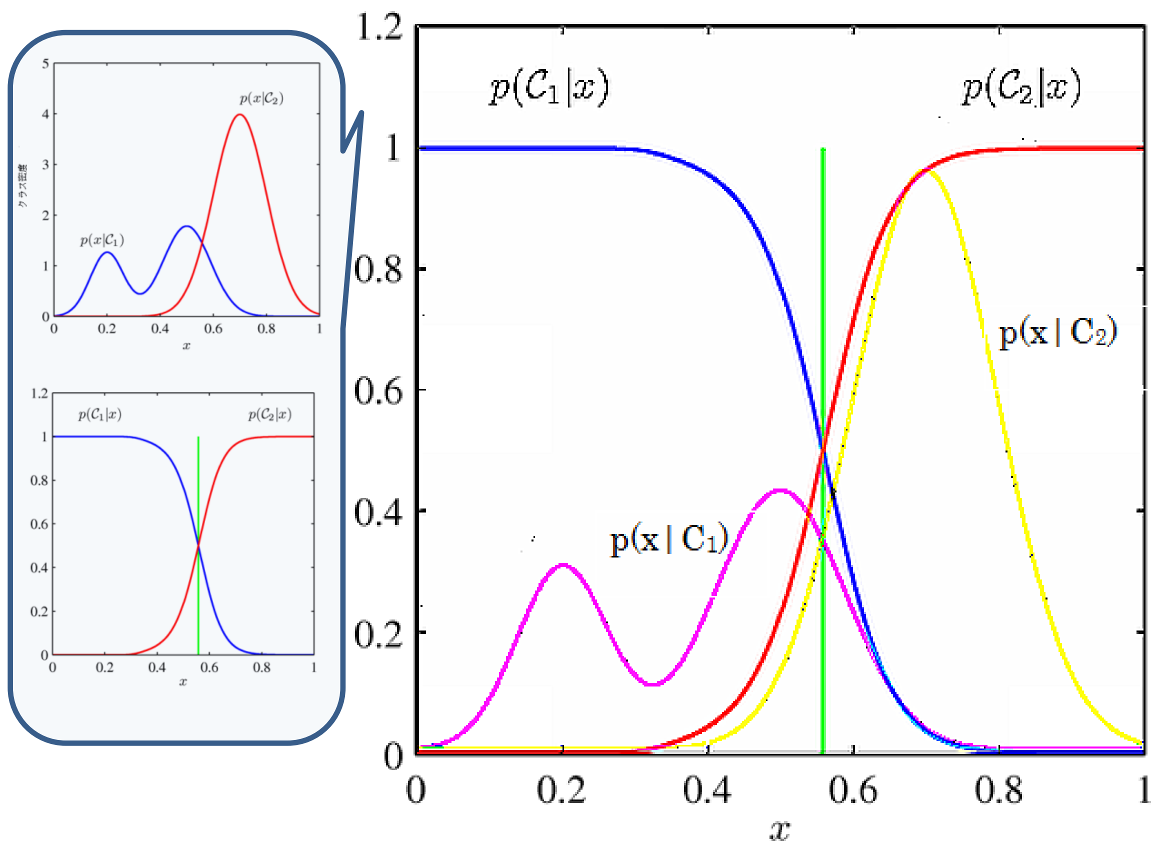
このモデル下で低い確率をとる新しいデータ点を見つけることに役立つ
→**外れ値検出**(outlier detection)/**新規性検出**(novelty detection)

アプローチ(b)の得失

- 計算資源の無駄がない(目的がクラス分類の決定のみ)

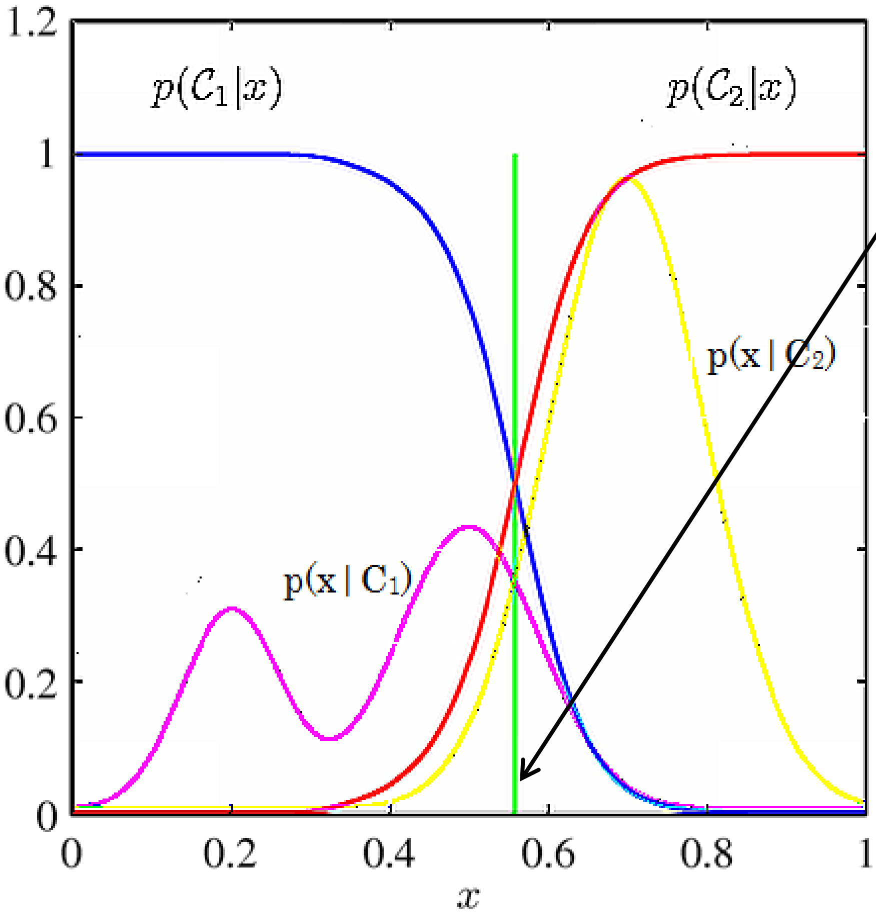
クラス分類の決定のみならば必要なのは事後確率p(Ck|x)

右の様にクラス条件付き分布は事後確率にあまり影響を及ぼさない複雑な構造をとり得る



アプローチ(c)の得失

- 推論と決定の段階は単一の学習問題に統合



このxを見つけることに相当
誤識別率最小の決定境界になる

しかしこの手法ではもはや
事後確率p(Ck|x)に接近できなくなる

事後確率を計算すべき理由

- リスク最小化

損失行列の要素が時間経過で変化する場合に
最小リスク決定基準の修正が容易

これの変更→ $\displaystyle\sum_k L_{kj}\, p(C_k|\mathbf{x})$

- 棄却オプション

棄却基準を定めることで、与えられた棄却データ点に対する
誤識別率や期待損失の最小化ができる

事後確率を計算すべき理由

- クラス事前確率の修正

例：X線の医療問題

訓練データ：一般の集団からのX線画像

・一般の集団で癌は稀
・癌のクラスが非常に小さい
・多様な癌のX線画像がない

要修正

良い訓練データ：各クラスのバランスが取れている

修正の補正が必要

$$\text{事後確率}\,(\text{修正後の訓練データ}) \times \frac{\text{元のクラスの割合}}{\text{修正後のクラスの割合}}$$

→最後に新しい事後確率が足して1になるように正規化

事後確率を計算すべき理由

- モデルの結合

例：X線画像の他に血液検査の情報も利用したい

1つの入力空間に纏めるより別々で扱う方が効率的
2つのモデルがそれぞれクラスの事後確率を持つ限り、
その出力を確率の規則に則り系統的に統合できる

x_I(X線画像)とx_B(血液データ)の分布が独立と仮定

$$p(\mathbf{x_I}, \mathbf{x_B}|C_k) = p(\mathbf{x_I}|C_k)p(\mathbf{x_B}|C_k)$$

分布をクラスC_kで条件付けしたとき独立し、条件付き独立の例となる

事後確率→
(要正規化)

$$\begin{aligned} p(C_k|\mathbf{x_I}, \mathbf{x_B}) &\propto p(\mathbf{x_I}, \mathbf{x_B}|C_k)p(C_k) \\ &\propto p(\mathbf{x_I}|C_k)p(\mathbf{x_B}|C_k)p(C_k) \\ &\propto \frac{p(C_k|\mathbf{x_I})p(C_k|\mathbf{x_B})}{p(C_k)} \end{aligned}$$