「パターン認識と機械学習」
1-2-3 ベイズ確率

新納浩幸

本章の内容

ベイズ確率とは何かという、概念的な説明

わかりずらいので、私なりのベイズ確率についての理解を話します

ベイズの定理

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

証明）

$$P(A,B) = P(A)P(B|A)$$

$$P(B,A) = P(B)P(A|B)$$

$P(A,B) = P(B,A)$　より明か ■

ベイズの定理の効果

X: 入力(問題)
c : 出力(答え)

$$\arg\max P(c \mid x) = \arg\max \frac{P(c)P(x \mid c)}{P(x)}$$

$$= \arg\max \underline{P(c)}\,\underline{P(x \mid c)}$$

事後確率

事前確率

尤度

問題に応じてこいつが利用できる

パラメータの捉え方

あるコインの表が出る確率 θ を知りたい

実験1）5 回投げたら 3回表 ⇨ $\theta = \frac{3}{5}$

実験2）500回投げたら 300回表 ⇨ $\theta = \frac{3}{5}$

頻度主義的な考え方だと同じ値だけど、何かちがう、
実験1の推定結果は怪しくて、実験2の推定結果は
信頼性が高い

実験3）5 回投げたら 5回表 ⇨ $\theta = 1$

Really ?

パラメータ自体が確率変数

θ は確率、だけどここでは 0 から 1 の間をとる変数と考える

確率変数

θ には分布がある

θ = 0.0 の確率は 0.00	θ = 0.6 の確率は 0.15
θ = 0.1 の確率は 0.01	θ = 0.7 の確率は 0.07
θ = 0.2 の確率は 0.02	θ = 0.8 の確率は 0.02
θ = 0.3 の確率は 0.07	θ = 0.9 の確率は 0.01
θ = 0.4 の確率は 0.15	θ = 1.0 の確率は 0.00
θ = 0.5 の確率は 0.50	

これは適当、、、だけど常識的な予想

ベイズ流のθの推定

arg max P(θ = x |実験1の結果)

= arg max P(θ = x) P(実験1の結果 | θ = x)

$$= \arg\max P(\theta = x) \cdot {}_5C_3 x^3 (1 - x)^2$$

X = 0.0 から 1.0 まで、
それぞれの値を代入して、
計算してみる

```
>prior <- c(0.00, 0.01, 0.02, 0.07, 0.15, 0.5, 0.15, 0.07, 0.02, 0.01, 0.00)
> x <- c(0:10)/10
> x
 [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

> dbinom(3,5,x)
 [1] 0.0000 0.0081 0.0512 0.1323 0.2304 0.3125 0.3456 0.3087 0.2048 0.0729
[11] 0.0000

> dbinom(3,5,x) * prior
 [1] 0.000000 0.000081 0.001024 0.009261 0.034560 0.156250 0.051840
0.021609
 [9] 0.004096 0.000729 0.000000

> prior[which.max(dbinom(3,5,x) * prior)]
[1] 0.5

> prior[which.max(dbinom(5,5,x) * prior)]
[1] 0.5

> prior[which.max(dbinom(7,7,x) * prior)]
[1] 0.07
```

R での解法

5回中5回でも
確率は 0.5

ベイズ流の考え方

ある問題があって、確率モデルでの解決を考える

このとき観測データから、モデルのパラメータを推定する問題になる

推定すべきパラメータを確率変数と考える

これだけが特徴ではないが、
利用する場面では、これが1つのポイント

ベイズの定理を利用すれば、
事前確率が利用できる

問題に応じた事前知識と考えて良い
事前知識を利用した解法・・・

2項分布の事前分布

パラメータ自体が確率変数

θ は確率、だけどここでは 0 から 1 の間をとる変数と考える

確率変数

θ には分布がある

θ = 0.0　の確率は　0.00
θ = 0.1　の確率は　0.01
θ = 0.2　の確率は　0.02
θ = 0.3　の確率は　0.07
θ = 0.4　の確率は　0.15
θ = 0.5　の確率は　0.50
θ = 0.6　の確率は　0.15
θ = 0.7　の確率は　0.07
θ = 0.8　の確率は　0.02
θ = 0.9　の確率は　0.01
θ = 1.0　の確率は　0.00

これは適当、、、だけど常識的な予想

適当すぎない？
連続値でもないし･･･

共役事前分布を使うのが計算が楽なので、
それを使う場合が多い、2項分布の共役事前分布は

ベータ分布

ベータ分布

$$f(x;a,b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$$

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$$

推定例

$$a = 5,\ b = 6$$

$$f(\theta) = P(\theta)P(D \mid \theta) = \frac{\theta^4(1-\theta)^5}{B(5,6)}\left({}_5C_3\theta^3(1-\theta)^2\right)$$

$$= \frac{{}_5C_3}{B(5,6)}\theta^7(1-\theta)^7$$

最大化

$$f'(\theta) = K\theta^6(1-2\theta) \quad \Rightarrow \quad \theta = \frac{1}{2}$$