Chap. 2

2.1 まで読み合わせ※

※ 十分統計量の話は p.113 2.4.1 でやろう。

2.1.1

p.67 のコイン投げの例のように、最尤法ではデータ数が少ないと過学習を起こすことがある。それを避けるために、ベイズ的なアプローチを用いる。

式変形は 自然科学の統計学 Chap.9 で既に学んでいるため、ここでは読み合わせ。

結論として、コインが表となる(ベイズ的)確率は

$$p(x=1|D) = \frac{m+a}{m+a+l+b}$$

で与えられる。$a, b$ が事前に入れ込まれているぶん、「我々の常識」から外れた確率になりづらい。

ここで $m, l \to \infty$ とすると、$m+l$ が $N$ (全データ数) に相当することになり

$$\frac{m+a}{m+a+l+b} \sim \frac{m}{m+l} = \frac{m}{N}$$

が得られる。これは最尤法の結果と一致する。つまり下記事実が成立する。

「最尤法の結果」 = 「ベイズアプローチによる結果」

ただし、データをたくさん取ってくるべし。

ちなみに、有限のデータ数の場合は、$\mu$ の事後平均は、事前平均と、$\mu$ の最尤推定量(学習データでつくったもの)の間の値になる。つまり

$$\mu_{post} = \lambda \mu_{pre} + (1-\lambda)\mu_{ML} \quad (0 \leq \lambda \leq 1)$$

となる。以下にこれが成立することを示す。(→ 対応:演習2.7)

$$\mu_{post} = \frac{m+a}{m+a+l+b} \quad (\because (2.20))$$

$$\mu_{pre} = \frac{a}{a+b} \quad (\because (2.15))$$

$$\mu_{ML} = \frac{m}{N} = \frac{m}{m+l} \quad (\because (2.81))$$

を利用する。

$$\frac{m+a}{m+a+l+b} = \lambda \frac{a}{a+b} + (1-\lambda)\frac{m}{m+l}$$

としたときに、$0 \leq \lambda \leq 1$ となる $\lambda$ が見つかればOK。実際初等計算によって

$$\lambda = \frac{a+b}{m+a+l+b}$$

$a, b, m, l > 0$ ゆえ $0 < \dfrac{a+b}{m+a+l+b} = \lambda < 1$ となり、

成立することが確かめられた。このことから、事後分布は事前分布と尤度の妥協点のような意味合いを持つと解釈できる。

次にベイズ学習の性質として、データを増やすほど事後分布の持つ不確実性が減る※ということが言えるか確かめる。

※分布のピークがはっきりあるということである。

結論から言うと、事後分布の分散は、事前分布の分散より小さくなる傾向があると言える※。これは常に（事後分布の分散）>（事前分布の分散）が成立するわけではなく、データセットによっては、この傾向に従わないことがあるという意味である。

※教科書内の「平均的に成立する」という表現が初見でさっぱりだったので、このような平易な表現に変えている。「平均的」の意味は、この後理解できるはず。

では、上記傾向があることを示す。つまり (2.24) を示す。（→対応；演習2.8）

$p(x, y)$ を2変数 $x, y$ の同時確率とする。

◎導入

(2.270) を示す。

$$(\text{RHS of } 2.270) = \int_y p(y) \int_x x\, p(x \mid y)\, dx\, dy$$

$$= \int_y p(y) \int_x x \frac{p(x, y)}{p(y)}\, dx\, dy \quad (\because \text{条件付確率のdef.})$$

$$= \int_x x \int_y p(x, y)\, dy\, dx$$

$$= \int_x x\, p(x)\, dx \quad (\because \text{周辺化})$$

$$= E(x) = (\text{LHS of } 2.270) \quad \blacksquare$$

導入の結果を使って(2.271)を示す.

$$(\text{RHS of } 2.271) = E_y\big(E_x(x^2|y) \underbrace{- E_x(x|y)^2\big) + E_y\big(E_x(x|y)^2}_{\text{キャンセル}}\big) - \big(E_y(E_x(x|y))\big)^2$$

$$= E_y(E_x(x^2|y)) - \big(E_y(E_x(x|y))\big)^2$$

$$= E_y(E_x(x^2|y)) - E(x)^2 \quad (\because (2.270);\ \text{導入の結果})$$

$$= \int_y p(y) \int_x x^2 p(x|y)\, dx\, dy - E(x)^2$$

$$= \int_x x^2 \int_y p(x|y)\, p(y)\, dy\, dx - E(x)^2$$

$$= \int_x x^2 \int_y \frac{p(x, y)}{\cancel{p(y)}} \cancel{p(y)}\, dy\, dx - E(x)^2$$

$$= \int x^2 p(x)\, dx - E(x)^2$$

$$= E(x^2) - E(x)^2 = (\text{LHS of } 2.271) \quad \blacksquare$$

$x$を$\theta$, $y$を$D$に読み変えてやると(2.24)が言える。