

## PRML4.5

2022 年 8 月 7 日

(問題)

(4.57) と (4.58) を使って、ガウス確率密度分布を用いた 2 クラス生成モデルにおけるクラスの事後確率に対する (4.65) の結果を導出せよ。また、パラメータ  $\mathbf{w}$  と  $w_0$  に対する結果 (4.66) と (4.67) を検証せよ。

(解答)

$$p(C_1|\mathbf{x}) = \sigma(a) \quad (4.57)$$

$$a = \ln \frac{p(\mathbf{x}|\mathbf{C}_1)\mathbf{p}(\mathbf{C}_1)}{p(\mathbf{x}|\mathbf{C}_2)\mathbf{p}(\mathbf{C}_2)} \quad (4.58)$$

$$p(\mathbf{x}|\mathbf{C}_k) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma^{-1}(\mathbf{x} - \mu_k) \right\} \quad (4.64)$$

これらの式を利用して (4.65) を導出する。

$$\begin{aligned} a &= \ln \frac{p(\mathbf{x}|\mathbf{C}_1)\mathbf{p}(\mathbf{C}_1)}{p(\mathbf{x}|\mathbf{C}_2)\mathbf{p}(\mathbf{C}_2)} \\ &= \ln \frac{p(\mathbf{x}|\mathbf{C}_1)}{p(\mathbf{x}|\mathbf{C}_2)} + \frac{\ln p(\mathbf{C}_1)}{\ln p(\mathbf{C}_2)} \\ &= \ln \frac{\exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu_1)^\top \Sigma^{-1}(\mathbf{x} - \mu_1) \right\}}{\exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu_2)^\top \Sigma^{-1}(\mathbf{x} - \mu_2) \right\}} + \frac{\ln p(\mathbf{C}_1)}{\ln p(\mathbf{C}_2)} \\ &= -\frac{1}{2}(\mathbf{x} - \mu_1)^\top \Sigma^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)^\top \Sigma^{-1}(\mathbf{x} - \mu_2) + \frac{\ln \mathbf{p}(\mathbf{C}_1)}{\ln \mathbf{p}(\mathbf{C}_2)} \\ &= \mu_1^\top \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_1^\top \Sigma^{-1} \mu_1 - \mu_2^\top \Sigma^{-1} \mathbf{x} + \frac{1}{2} \mu_2^\top \Sigma^{-1} \mu_2 + \frac{\ln \mathbf{p}(\mathbf{C}_1)}{\ln \mathbf{p}(\mathbf{C}_2)} \\ &= (\mu_1 - \mu_2)^\top \Sigma^{-1} \mathbf{x} - \frac{1}{2} \mu_1^\top \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_2^\top \Sigma^{-1} \mu_2 + \frac{\ln \mathbf{p}(\mathbf{C}_1)}{\ln \mathbf{p}(\mathbf{C}_2)} \\ &= (\Sigma^{-1}(\mu_1 - \mu_2))^\top \mathbf{x} - \frac{1}{2} \mu_1^\top \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_2^\top \Sigma^{-1} \mu_2 + \frac{\ln \mathbf{p}(\mathbf{C}_1)}{\ln \mathbf{p}(\mathbf{C}_2)} (\because (\Sigma^{-1})^\top = \Sigma^{-1}) \end{aligned}$$

よって、

$$\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_2) \quad (4.66)$$

$$w_0 = -\frac{1}{2}\mu_1^\top \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^\top \Sigma^{-1}\mu_2 + \frac{\ln p(\mathbf{C}_1)}{\ln p(\mathbf{C}_2)} \quad (4.67)$$

と定義すれば

$$p(C_1|\mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + \mathbf{w}_0) \quad (4.65)$$

が導ける。

(補足: 分類問題におけるアプローチについて)

分類問題では以下の 3 つのアプローチが考えられる

(a) 生成モデル

生成モデルはデータが生成される分布を推定し、ベイズの定理からその分布を用いてクラス事後確率を求める。クラス事後確率を求めたら、決定理論により新たな入力  $\mathbf{x}$  のクラス属性を決めることが出来る。すなわち分類対象となるデータがどのような確率モデルから生成されたかをモデル化し、そのモデルに基づいて分類を行う手法。

例: ナイーブベイズモデル

(b) 識別モデル

識別モデルは最初にクラス事後確率を決める推論問題を解き、次に決定理論により新たな入力  $\mathbf{x}$  のクラス属性を決める。クラス識別モデルでは事後確率を直接モデル化している。

例: ロジスティック回帰、ソフトマックス交差エントロピーを使ったニューラルネットワーク (c) 識別関数  
各入力  $\mathbf{x}$  から直接クラスラベルを割り当てる関数  $f(\mathbf{x})$  を見つける

例: SVM

生成モデルを使う利点としてはデータの周辺分布を求めることができて、それを外れ値検出や、人口データの生成に利用することが出来る。一方で識別モデルと比較して計算コストが高いというデメリットもある。計算コストの観点でいうと確率分布の推定を行わない識別関数が最も低い、出力に対して確率的な解釈が得られないため、外れ値検出といったようなことができない。

まとめると

計算コスト: 生成モデル > 識別モデル > 識別関数

解釈性: 識別関数 > 識別モデル > 生成モデル

どのアプローチを採択するのが最適かはデータに依る。