第4章
最尤推定法

4.1 最尤推定法の定義
4.2 ガウスモデル

07T4072F　岡崎駿

パラメトリック法（前回）

・パラメトリック法

パラメトリックと呼ばれる有限個のパラメータで記述された確率密度関数の集合の中から、真の確率密度関数を最もよく近似するものを選ぶ。

『最尤推定法』

パラメトリックモデル

有限個のパラメータで記述された確率密度関数の集合

・パラメトリックモデル

$$q(\mathbf{x};\boldsymbol{\theta})$$

・確率変数: $\mathbf{x}$

・パラメータ: $\boldsymbol{\theta} = (\theta^{(1)},\ldots,\theta^{(t)})^{\mathrm{T}}$

・パラメータの定義域: Θ

尤度と最尤推定法(1)

目的:確率密度関数を近似したい

手元の訓練標本が最も生起しやすいように
パラメータを決定する

尤度と最尤推定法（2）

・尤度

訓練標本 $\{\mathbf{x}\}_{i=1}^{n}$ が最も生起しやすい確率を、
パラメータ $\boldsymbol{\theta}$ の関数と見たもの

$$L(\boldsymbol{\theta}) := \prod_{i=1}^{n} q(\mathbf{x}_i; \boldsymbol{\theta})$$

・最尤推定法

尤度が最大になるようにパラメータ値を決定する方法

$$\hat{\boldsymbol{\theta}}_{ML} := \underset{\boldsymbol{\theta} \in \Theta}{\arg\max}\, L(\boldsymbol{\theta})$$

$\hat{\boldsymbol{\theta}}_{ML}$ ：最尤推定法によって得られるパラメータ値

確率密度関数と尤度方程式

パラメータ $\hat{\boldsymbol{\theta}}_{ML}$ を用いた確率密度関数

$$\hat{p}(\mathbf{x}) = q(\mathbf{x}; \hat{\boldsymbol{\theta}}_{ML})$$

パラメトリックモデル $q(\mathbf{x};\boldsymbol{\theta})$ が $\boldsymbol{\theta}$ に関して微分可能なとき

$$\left.\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = \mathbf{0}_t$$

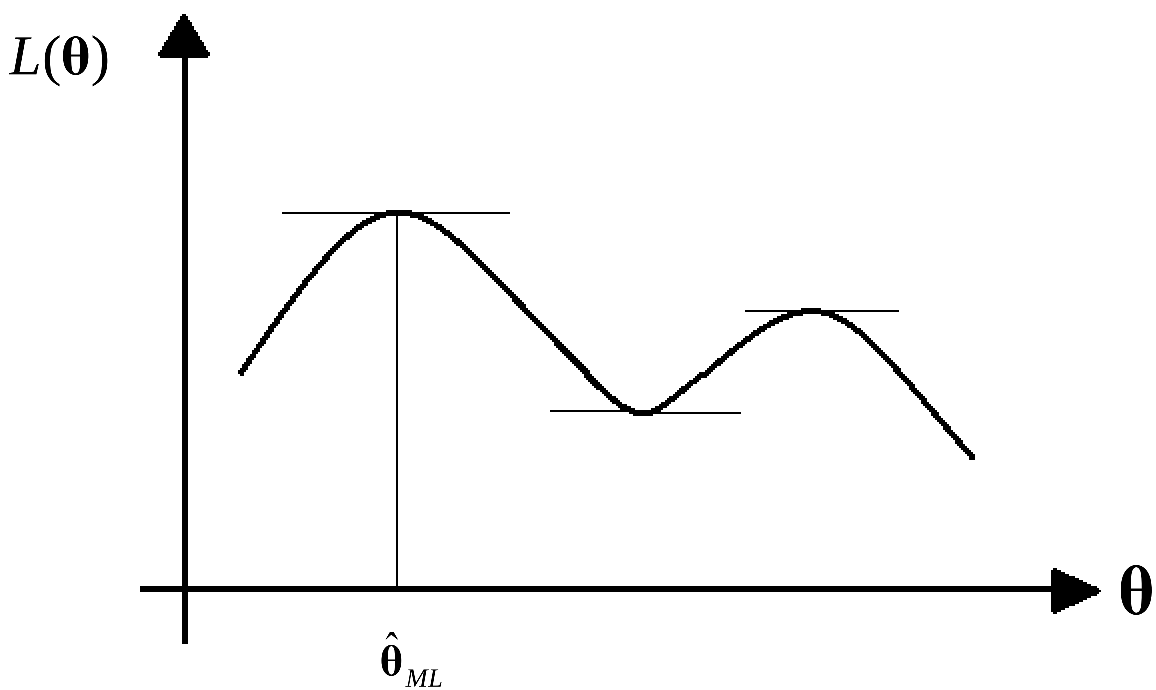
『尤度方程式』

尤度方程式と最尤推定解

『尤度方程式』 → 『最尤推定解』

◯必要条件

×十分条件



対数尤度

「尤度を最大にするパラメータ」

||

「対数をとった尤度を最大にするパラメータ」

※具体的な計算方法は用いるパラメトリックモデルによって異なる！

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max} \log L(\boldsymbol{\theta})$$

$$= \underset{\boldsymbol{\theta}\in\Theta}{\arg\max}\left[\sum_{i=1}^{n} \log q(\mathbf{x}_i;\boldsymbol{\theta})\right]$$

対数尤度

$$\left.\frac{\partial}{\partial\boldsymbol{\theta}} \log L(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = \mathbf{0}_t$$

尤度方程式

ガウスモデル

ガウス分布をパラメトリックモデルとして用いたもの

d 次元ベクトルのパターン $\mathbf{x}$ に対する一般形

$$q(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}}\det(\boldsymbol{\Sigma})^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

- d 次元期待値ベクトル : $\boldsymbol{\mu}$
- $d\times d$ 行列のガウスモデルのパラメータ（分散共分散行列）: $\boldsymbol{\Sigma}$

※$\boldsymbol{\Sigma}$ は正定値行列かつ対称行列の必要がある。
そのため任意の非ゼロベクトル φ に対して $\varphi^{\mathrm{T}}\boldsymbol{\Sigma}\varphi > 0$ を満たす。

ガウスモデル-尤度方程式-

ガウスモデルの特徴

→パラメータ $\boldsymbol{\mu}$と $\boldsymbol{\Sigma}$がそれぞれ分布の期待値と分散共分散行列に対応している。(2.1節24p)

$$\begin{cases} \boldsymbol{\mu} = E_{\mathbf{x}}[\mathbf{x}] = \int \mathbf{x} q(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} \\ \boldsymbol{\Sigma} = V_{\mathbf{x}}[\mathbf{x}] = \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} q(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} \end{cases}$$

ガウスモデルの尤度方程式は

$$\begin{cases} \left. \dfrac{\partial}{\partial \boldsymbol{\mu}} \log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \right|_{\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}_{ML}} = \mathbf{0}_d \\ \left. \dfrac{\partial}{\partial \boldsymbol{\Sigma}} \log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \right|_{\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}_{ML}, \boldsymbol{\mu} = \hat{\boldsymbol{\mu}}_{ML}} = \mathbf{0}_{d \times d} \end{cases}$$

ガウスモデル-最尤推定法-

前ページの尤度方程式を解けば $\boldsymbol{\mu}$ 、$\boldsymbol{\Sigma}$ の最尤推定量 $\hat{\boldsymbol{\mu}}_{ML}$ 、$\hat{\boldsymbol{\Sigma}}_{ML}$ が得られる。

$$
\begin{cases}
\hat{\boldsymbol{\mu}}_{ML} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbf{x}_i \\
\hat{\boldsymbol{\Sigma}}_{ML} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{ML})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{ML})^{\mathrm{T}}
\end{cases}
$$

真の期待値を標本平均、
真の分散共分散行列を標本分散共分散行列
で表している。

共分散のないガウスモデル(1)

- 一般の分散共分散行列 Σ のi番目の対角成分を $(\sigma^{(i)})^2$ で表す。

$$q(\mathbf{x};\boldsymbol{\mu},\sigma^{(1)},\ldots,\sigma^{(d)})=\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi(\sigma^{(j)})^2}}\exp\left(-\frac{(x^{(j)}-\mu^{(j)})^2}{2(\sigma^{(j)})^2}\right)$$

- $\sigma^{(j)}$ の最尤推定量

$$\hat{\sigma}^{(j)}{}_{ML}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^{(j)}-\mu_i^{(j)})^2}$$

共分散のないガウスモデル（2）

・全ての分散 $(\sigma^{(j)})^2$ が全て等しいガウスモデル
  ※共通の分散を σ で表す。

$$q(\mathbf{x};\boldsymbol{\mu},\sigma)=\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu})}{2\sigma^2}\right)$$

・σ の最尤推定量

$$\hat{\sigma}_{ML}=\sqrt{\frac{1}{nd}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathrm{T}}(\mathbf{x}_i-\boldsymbol{\mu})}=\sqrt{\frac{1}{d}\sum_{j=1}^{d}(\hat{\sigma}_{ML}^{(j)})^2}$$