練習問題 6-1

(1)

ノイズ（誤差項）$\boldsymbol{\varepsilon}$ が正規分布 $N(0,1)$ に従うから，最尤推定による解と最小二乗法による解は一致する。scipy.optimize.leastsq を用いて最小二乗法による解を求めると，

$$a = 1.059 \tag{1}$$

と求まる。

(2)

a の事前分布が $p(a) = N(1, 0.09)$ であるとする。ノイズの分散を $\sigma_{\mathrm{e}}{}^2$，a の平均を μ_a，分散を $\sigma_a{}^2$ とする。まず一般の線形回帰問題を考える。以下のようなモデルの MAP 推定を行う。

$$y = \boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{X} + \boldsymbol{\varepsilon} \tag{2}$$

$$\varepsilon \sim N(0, \sigma_{\mathrm{e}}{}^2) \tag{3}$$

$$p(\boldsymbol{a}) = N(\mu_a, \sigma_a{}^2) \tag{4}$$

対数尤度関数は以下のようになる。

$$\begin{aligned}
LL(\boldsymbol{a}, \sigma_{\mathrm{e}}{}^2) &= \log \prod_{i=1}^{N} p(y_i|\boldsymbol{x}_i, \boldsymbol{a}, \sigma_{\mathrm{e}}{}^2) p(\boldsymbol{a}) \\
&= \sum_{i=1}^{N} \log \left\{ p(y_i|\boldsymbol{x}_i, \boldsymbol{a}, \sigma_{\mathrm{e}}{}^2) p(\boldsymbol{a}) \right\} \\
&= \sum_{i=1}^{N} \log \left\{ \frac{1}{\sqrt{2\pi\sigma_{\mathrm{e}}{}^2}} \exp\left(-\frac{(\varepsilon_i)^2}{2\sigma_{\mathrm{e}}{}^2} \right) \cdot \frac{1}{\sqrt{2\pi\sigma_a{}^2}} \exp\left(-\frac{(a_i - \mu_a)^2}{2\sigma_a{}^2} \right) \right\} \\
&= \sum_{i=1}^{N} \left\{ -\frac{1}{2\sigma_{\mathrm{e}}{}^2} \varepsilon_i{}^2 - \frac{1}{2\sigma_a{}^2} (a_i - \mu_a)^2 \right\} - \frac{N}{2} \left\{ \log\left(2\pi\sigma_{\mathrm{e}}{}^2\right) + \log\left(2\pi\sigma_a{}^2\right) \right\} \\
&= -\frac{1}{2\sigma_{\mathrm{e}}{}^2} \left(\|\boldsymbol{\varepsilon}\|^2 + \frac{\sigma_{\mathrm{e}}{}^2}{\sigma_a{}^2} \|(\boldsymbol{a} - \mu_a)\|^2 \right) - N \log(2\pi\sigma_{\mathrm{e}}\sigma_a) \\
&= -\frac{1}{2\sigma_{\mathrm{e}}{}^2} \left(\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}\|^2 + \frac{\sigma_{\mathrm{e}}{}^2}{\sigma_a{}^2} \|(\boldsymbol{a} - \mu_a)\|^2 \right) - N \log(2\pi\sigma_{\mathrm{e}}\sigma_a)
\end{aligned}$$

よって，

$$LL(\boldsymbol{a}, \sigma_{\mathrm{e}}{}^2) \text{ が最大} \quad \Leftrightarrow \quad \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}\|^2 + (\sigma_{\mathrm{e}}{}^2/\sigma_a{}^2)\|(\boldsymbol{a} - \mu_a)\|^2 \text{が最小} \tag{5}$$

ここで，

$$\lambda = \frac{\sigma_{\mathrm{e}}{}^2}{\sigma_a{}^2} \tag{6}$$

とおくと，$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}\|^2 + \lambda\|(\boldsymbol{a} - \mu_a)\|^2$ を最小にすればよいことがわかる。

　ここで，配布資料より，

$$\begin{aligned}\log p(\boldsymbol{a}|\boldsymbol{X},\boldsymbol{y},\sigma^2) &= -\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a}\|^2 - \frac{\lambda}{2\sigma^2}\|\boldsymbol{a}\|^2 + \text{Const.} \\ &= -\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a}\|^2 + \lambda\|\boldsymbol{a}\|^2\right) + \text{Const.}\end{aligned} \tag{7}$$

を最大にする $\boldsymbol{a}$ は，

$$\hat{\boldsymbol{a}}_{\text{MAP}} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{y}\right) \tag{8}$$

となることがわかる。よって，

$$\boldsymbol{b} = \boldsymbol{a} - \mu_a, \quad \boldsymbol{z} = \boldsymbol{y} - \mu_a\boldsymbol{X} \tag{9}$$

と変換すると，

$$\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a}\|^2 + \lambda\|(\boldsymbol{a}-\mu_a)\|^2 = \|\boldsymbol{z}-\boldsymbol{X}\boldsymbol{b}\|^2 + \lambda\|\boldsymbol{b}\| \tag{10}$$

を最小にすればよいから，その解は，

$$\hat{\boldsymbol{b}}_{\text{MAP}} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{z}\right) \tag{11}$$

結局，対数尤度関数 $LL(\boldsymbol{a},{\sigma_\text{e}}^2)$ を最大にする $\boldsymbol{a}$ は，

$$\begin{aligned}\hat{\boldsymbol{a}}_{\text{MAP}} &= \hat{\boldsymbol{b}}_{\text{MAP}} + \mu_a \\ &= \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{z}\right) + \mu_a \\ &= \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^T\boldsymbol{y} - \mu_a\boldsymbol{X}^T\boldsymbol{X}\right) + \mu_a\end{aligned} \tag{12}$$

と表せる。

　式 (12) に，

$$\mu_a = 1 \tag{13}$$

$$\lambda = \frac{{\sigma_\text{e}}^2}{{\sigma_a}^2} = \frac{1}{0.09} = \frac{100}{9} \tag{14}$$

$$\tag{15}$$

と $\boldsymbol{X}$, $\boldsymbol{y}$ を代入すると，

$$a = 1.028 \tag{16}$$

を得る。

　以下に MLE，MAP 推定それぞれの結果のグラフを示した。

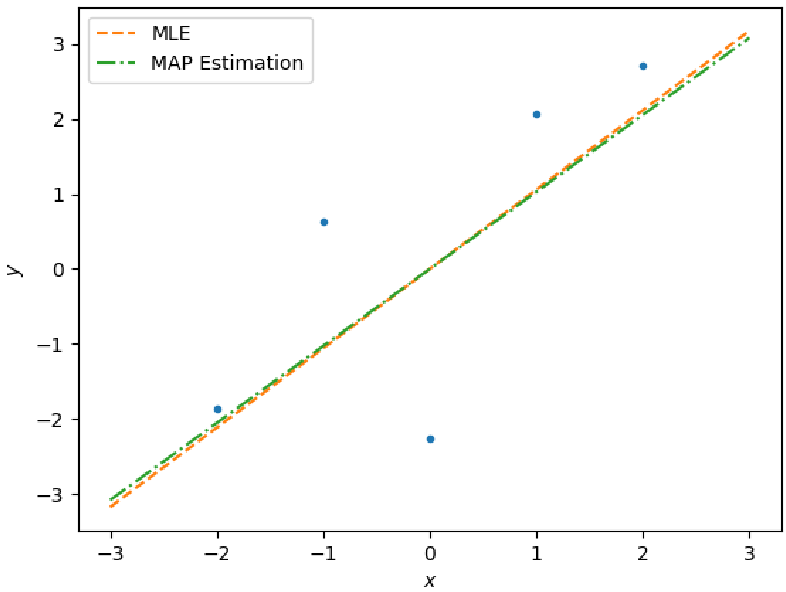

図 1: 【練習問題 6-1】 MLE と MAP 推定による線形回帰のグラフ