PRML 1 章 演習問題

mei

問題 1.1

$$y(x, \boldsymbol{w}) = \sum_{j=i}^{M} w_j x_j, \qquad E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \boldsymbol{w}) - t_n\}^2 \tag{1}$$

二乗和誤差 (1) を最小にする係数 $\boldsymbol{w} = \{w_i\}$ が次の線形方程式の解として与えられることを示す

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{2}$$

ただし、

$$A_{ij} = \sum_{n=1}^{N}(x_n)^{i+j}, \qquad T_i = \sum_{n=1}^{N}(x_n)^i t_n$$

解 答

$$\begin{aligned}
\frac{\partial E}{\partial w_i} &= \sum_{n=1}^{N}(y_n(x_n, \boldsymbol{w}) - t_n)\frac{\partial}{\partial w_i}(y_n(x_n, \boldsymbol{w}) - t_n) \\
&= \sum_{n=1}^{N}\left(\sum_{j=0}^{M}(x_n)^j w_j - t_n\right)(x_n)^i \\
&= 0
\end{aligned}$$

の時に最小になる

よって、

$$\begin{gathered}
\sum_{n=1}^{N}\left(\sum_{j=0}^{M}(x_n)^j w_j\right)(x_n)^i = \sum_{n=1}^{N}(x_n)^i t_n \\
\sum_{j=0}^{M}\sum_{n=1}^{N}(x_n)^{i+j} w_j = \sum_{n=1}^{N}(x_n)^i t_n
\end{gathered}$$

となり、(2) を得る。

補 足

E は w_i について下に凸な二次関数だから $\frac{\partial E}{\partial w_i} = 0$ となる w_i が E を最小にする

問題 1.2

式 (1) に正則化項がついた次の誤算関数を最小にする係数 w_i が満たす式を求める

$$\widetilde{E}(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \boldsymbol{w}) - t_n\}^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|^2 \tag{3}$$

解 答

問題 1.1 と同様の計算により、

$$\begin{aligned}\frac{\partial E}{\partial w_i} &= \sum_{n=1}^{N}\left(\sum_{j=0}^{M}(x_n)^j w_j - t_n\right)(x_n)^i + \lambda w_i \\ &= 0\end{aligned}$$

よって、

$$\sum_{j=0}^{M} A_{ij} w_j = T_i - \lambda w_i \tag{4}$$

ただし、

$$A_{ij} = \sum_{n=1}^{N}(x_n)^{i+j}, \qquad T_i = \sum_{n=1}^{N}(x_n)^i t_n$$

が w_i が満たすべき線形方程式系になる

問題 1.3

下の表のような箱がある。箱を $p(r)=0.2, p(b)=0.2, p(g)=0.6$ の確率でランダムに選び果物を区別せず等確率で取り出す時、(i) りんごを取り出す確率、(ii) 選んだ果物がオレンジであった時それが g の箱から取り出されたものである確率、を求める

箱	りんご	オレンジ	ライム
r	3	4	3
b	1	1	0
g	3	3	4

解答

(i)

$$\begin{aligned} p(F=a) &= \sum_B p(F=a|B)p(B) \\ &= p(F=a|B=r)p(B=r) + p(F=a|B=b)p(B=b) + p(F=a|B=g)p(B=g) \\ &= \frac{3}{10}\cdot\frac{1}{5} + \frac{1}{2}\cdot\frac{1}{5} + \frac{3}{10}\cdot\frac{3}{5} \\ &= \frac{17}{50} \end{aligned}$$

(ii)

$$\begin{aligned} p(B=g|F=o) &= \frac{p(F=o|B=g)p(B=g)}{p(F=o)} \\ &= \frac{\frac{3}{10}\cdot\frac{3}{5}}{\frac{4}{10}\cdot\frac{1}{5}+\frac{1}{2}\cdot\frac{1}{5}+\frac{3}{10}\cdot\frac{3}{5}} \\ &= \frac{1}{2} \end{aligned}$$

問題 1.4

連続変数 x 上で定義された確立密度 $p_x(x)$ を考える。$x = g(y)$ による変換を施した場合 y についての密度を最大にする位置 $\widehat{y}$ と x についての密度を最大にする位置 $\widehat{x}$ との間に一般には $\widehat{x} = g(\widehat{y})$ の関係が成り立たないことを示す。また、線形変換ではこの式が成り立つことも示す。

解 答

問題 1.5

次の式を示す。

$$\mathrm{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

解 答

$$
\begin{aligned}
\mathrm{var}[f] &= \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] \\
&= \int p(x)(f(x) - \mathbb{E}[f(x)])^2 dx \\
&= \int p(x)f(x)^2 dx - 2\mathbb{E}[f(x)] \int p(x)f(x)dx + \mathbb{E}[f(x)]^2 \int p(x)dx \\
&= \mathbb{E}[f(x)^2] - 2\mathbb{E}[f(x)]\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2 \\
&= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2
\end{aligned}
$$

より示せた。

補 足

3 行目から 4 行目では $E[f(x)]$ の定義と $p(x)$ が規格化されていることを用いた。

問題 1.6

2 つの変数 x, y が独立ならそれらの共分散が 0 になることを示す

解 答

$$\mathrm{cov}[x, y] = \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \tag{5}$$

一方、

$$\mathbb{E}_{x,y}[xy] = \sum_x \sum_y p(x, y)xy$$

であり、x, y が独立であることから $p(x, y) = p(x)p(y)$ とかけるので、

$$\begin{aligned}\mathbb{E}_{x,y}[xy] &= \sum_x \sum_y p(x, y)xy \\ &= \sum_x \sum_y p(x)p(y)xy \\ &= \sum_x p(x)x \sum_y p(y)y \\ &= \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

となる。
これと (5) 式より、

$$\mathrm{cov}[x, y] = 0$$

が示される。

問題 1.7

次の 1 変数ガウス分布に関する規格化条件を証明する

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)dx = 1$$

解 答

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)dx = I$$

とおき、I^2 を考える。

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)dx \int_{-\infty}^{\infty} \mathcal{N}(y|\mu, \sigma^2)dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|\mu, \sigma^2)dxdy \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) dxdy \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left[(x-\mu)^2 + (y-\mu)^2\right]\right) dxdy \\
&= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2+y^2)\right) dxdy
\end{aligned}$$

最後の行は改めて $x - \mu \to x, y - \mu \to y$ とおいた。

ここで、極座標変換を施すと、

$$\begin{aligned}
I^2 &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2+y^2)\right) dxdy \\
&= \frac{1}{2\pi\sigma^2} \int_{0}^{\infty} \int_{0}^{2\pi} \exp\left(-\frac{1}{2\sigma^2}r^2\right) rdrd\theta \\
&= \frac{1}{\sigma^2} \int_{0}^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) rdr \\
&= -\int_{0}^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) \left(-\frac{1}{2\sigma^2}r^2\right)' dr \\
&= -\left[\exp\left(-\frac{1}{2\sigma^2}r^2\right)\right]_0^{\infty} \\
&= 1
\end{aligned}$$

よって、$I \geq 0$ より $I = 1$ が示された

問題 1.8

1 変数ガウス分布の平均値が μ で与えられることを示し、次に規格化条件

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1$$

の両辺を σ^2 に関して微分することで、ガウス分布の 2 次のモーメントを求める。最後に分散を求める。

まず、$\mathbb{E}[x]$ を求める。

$$\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x dx \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) x dx
\end{aligned}$$

ここで、$x - \mu \to x$ と置き換えると、

$$\begin{aligned}
\mathbb{E}[x] &= \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right)(x+\mu) dx \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) x dx + \mu \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx \\
&= C \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right)\left(-\frac{1}{2\sigma^2}x^2\right)' dx + \mu \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx \qquad (C : \mathrm{Const.}) \\
&= 0 + \mu \cdot 1 \\
&= \mu
\end{aligned}$$

$\mathbb{E}[x^2]$ については、

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx &= \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\
&= 1
\end{aligned}$$

の両辺を σ^2 について微分すると、

$$\begin{aligned}
&-\frac{1}{2}\frac{1}{(2\pi)^{1/2}}\frac{1}{\sigma^3} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx + \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)\left(\frac{1}{2\sigma^4}(x-\mu)^2\right) dx \\
&= -\frac{1}{2\sigma^2}\frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx + \frac{1}{2\sigma^4}\frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)(x-\mu)^2 dx \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mu^2) \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbb{E}[x^2] - \mu^2) \\
&= 0
\end{aligned}$$

よって、

$$\mathbb{E}[x^2] = \mu^2 + \sigma^2$$

を得る。以上より、分散は

$$\begin{aligned}
\mathrm{var}[x] &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\
&= \mu^2
\end{aligned}$$

と求められる。

問題 1.9

1 変数ガウス分布 (6) のモードおよび多変量ガウス分布 (7) のモードが μ や $\boldsymbol{\mu}$ で与えられることを示す。

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \tag{6}$$

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{7}$$

解答

まず 1 変数について考える。上に凸なので微分が 0 になる点で確率分布が最大値を取る

$$\begin{aligned}
\frac{d}{dx}\mathcal{N}(x|\mu,\sigma^2) &= \frac{d}{dx}\left[\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)\right] \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} \frac{-1}{\sigma^2}(x-\mu) \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\
&= 0
\end{aligned}$$

よって $x=\mu$ で $\mathcal{N}$ が最大になる、つまりモードが μ で与えられる。
次に多変量の場合を考える。指数部分の二次形式を考えると、

$$-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = -\frac{1}{2}\sum_{j=1}^{D}\sum_{i=1}^{D}(x_i-\mu_i)^2\mathrm{cov}[x_i,x_j]$$

となるため、勾配が 0 になるとき

$$\begin{aligned}
\frac{\partial}{\partial x_k}\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{-1}{2}\sum_{i=1}^{D}2(x_k-\mu_k)\mathrm{cov}[x_i,x_k] \\
&= -(x_k-\mu_k)\sum_{i=1}^{D}\mathrm{cov}[x_i,x_k]
\end{aligned}$$

よって、$x_k=\mu_k$ の時 $\mathcal{N}$ が最大になる、つまりモードが $\boldsymbol{\mu}$ で与えられる

問題 1.10

2 変数が独立の時、それらの和の平均と分散が

$$\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$$
$$\mathrm{var}[x+z] = \mathrm{var}[x] + \mathrm{var}[z]$$