PRML Chapter 6 Answer

6.3

$$\|\mathbf{x}-\mathbf{x}_n\|^2 = (\mathbf{x}-\mathbf{x}_n)^T(\mathbf{x}-\mathbf{x}_n)$$

$$= \phi(\mathbf{x}_n)^T\phi(\mathbf{x}_n) \quad (\phi(x_n) = x - x_n)$$

$$= k(\mathbf{x}_n, \mathbf{x}_n)$$ とできる.

6.4

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 7 \end{pmatrix}$$ のとき.

Eigen value of A is 2, 7

6.5 (6.13) : $k(x,x') = ck_1(x,x')$, $c>0$

(k_1 is a kernel function)

$$ck_1(x,x') = c\phi(x)^T\phi(x')$$
$$= \underbrace{\sqrt{c}\,\phi(x)^T\sqrt{c}\,\phi(x')}_{\text{kernel function の形}}$$

6.6 (6.15) $k(x,x') = g(k_1(x,x'))$ (k_1 is also a kernel function)

(g is polynomial)

$$g(x) = \sum_i a_i x^i \Rightarrow g(k_1) = \sum_i a_i k_1^i$$

(6.18) : $k(x,x') = k_m(x,x')\cdot k_n(x,x')$

(k_m, k_n are kernel functions)

$\Rightarrow k(x,x')$ is a kernel function

を用いて帰納的に

k_1^i も kernel function

(6.13) より、定数倍も kernel function なので:

$\sum_i a_i k^i$ is a kernel function ㊄

(6.16)

use Taylor Expansion of the exponential function

6.7 (6.17) k_1, k_2 の Gram Matrix を K_1, K_2 とする。

$K = K_1 + K_2$

$\omega^T K_1 \omega \geqq 0$, $\omega^T K_2 \omega \geqq 0$ より. $\omega^T (K_1 + K_2) \omega$

$= \omega^T K \omega \geqq 0$ ㊄

(6.18) $k(x, x') = k_1(x, x') k_2(x, x')$

$= \phi^T(x) \phi(x') \psi(x) \psi(x')$

$= \sum_{i=1}^{M} \phi_i(x) \phi_i(x') \sum_{j=1}^{N} \psi_j(x) \psi_j(x')$

$$= \underbrace{\sum_{i=1}^{M} \sum_{j=1}^{N} \phi_i(x) \psi_j(x) \phi_i(x') \psi_j(x')}$$

kernel function の形になる。

6.8 (6.19) $k(x, x') = k_3(\phi(x), \phi(x'))$

$$\to \psi(\phi(x))^T \psi(\phi(x')) = \underbrace{\Phi(x)^T \Phi(x')}$$

kernel function.

(6.20) $k(x, x') = x^T A x$

Aは対称な半正定値行列より、対角化して、

$A = U^T E U$, $x^T A x = x^T U^T E U x =$

$(Ux)^T = x^T U^T$ より、 $= \underbrace{(Ux)^T E (Ux)}$

kernel functionの

linear combination.

6.10 $y(x) = k(x)^T a$

$$= \sum_{n=1}^{N} f(x_n) f(x) \cdot a_n \propto f(x)$$ 図

6.11 $\exp(x^T x'/\sigma^2) = \underbrace{\sum_{n=0}^{\infty} \frac{(x^T x'/\sigma^2)^n}{n}}$ ← 有効なkernel.

よって、(6.23)も有効なkernel.

6.13 $k(x,x') = g(\theta,x)^T F^{-1} g(\theta,x)$ Fisher kernel

~~微分~~演算子~~行列~~A(大きさ1)が $A^TA=I$ となって、

$g(\psi(\theta),x)^T F^{-1} g(\psi(\theta),x)$ 同じ

6.14. $p(x|\mu) = \mathcal{N}(x|\mu,S)$

$g(\mu,x) = \nabla_\mu \ln p(x|\mu) = S^{-1}(x-\mu)$

according to (6.34), $F = \mathbb{E}_x[S^{-1}(x-\mu)(x-\mu)^T S^{-1}]$ $\mathbb{E}_x[(x-\mu)^T(x-\mu)] = S$ より

$F = S^{-1}$. $k(x,x') = (x-\mu)^T S^{-1}(x'-\mu)$

6.17 $E[y(x)] = \frac{1}{2}\sum_{n=1}^{N} \int \{y(x_n+\xi) - t_n\}^2 \nu(\xi)\, d\xi$

$E[y(x) + \varepsilon\eta(x)] = E[y(x)] + \varepsilon\sum \int \eta(x_n+\xi)\{y(x_n+\xi)-t_n\}\nu(\xi)\, d\xi + O(\varepsilon^2)$

よって. $\sum \int \eta(x_n+\xi)\{y(x_n+\xi)-t_n\}\nu(\xi)\, d\xi = 0$ …①

$\eta(x_n) = \delta(x_n - x)$ とする

$\int \delta(x_n+\xi-x) \cdot \{y(x_n+\xi) - t_n\}\nu(\xi)\, d\xi$

$= \{y(x) - t_n\} \cdot \nu(-x_n + x)$ より.

① $= \sum \{y(x) - t_n\} \cdot \nu(-x_n + x) = 0$

よって、

$$y(\mathbf{x}) = \frac{\sum t_n \nu(\mathbf{x}-\mathbf{x}_n)}{\sum \nu(\mathbf{x}-\mathbf{x}_n)}.$$

$\boxed{\text{Q}}$

6.18 $f(x,t) = f(\mathbf{z}) = \mathcal{N}(\mathbf{z}|0,\sigma^2) = \frac{1}{2\pi}\cdot\frac{1}{\sigma^2}\exp\left\{-\frac{1}{2}\mathbf{z}\ \frac{1}{\sigma^2}\mathbf{I}\mathbf{z}\right\}$

(6.47)より、

$$g(x) = \int_{-\infty}^{\infty} f(x,t) = \int_{-\infty}^{\infty}\frac{1}{2\pi}\cdot\frac{1}{\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(x^2+t^2)\right\}dt$$

$$= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2\sigma^2}\right)dt$$

$$= \frac{1}{2\pi\sigma^2}\cdot\sqrt{2\pi\sigma^2}\cdot\exp\left(-\frac{x^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$$

(6.46)より、

$$k(\mathbf{x},\mathbf{x}_n) = \frac{\exp\left\{-\frac{(x-x_n)^2}{2\sigma^2}\right\}}{\sum_m \exp\left(-\frac{(x-x_m)^2}{2\sigma^2}\right)}$$

(6.48)についても同様で、

$$p(t|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{\sum_n \exp\left\{-\frac{(x-x_n)^2+(t-t_n)^2}{2\sigma^2}\right\}}{\sum_m \exp\left\{-\frac{(x-x_m)^2}{2\sigma^2}\right\}}$$

$$= \sum_n k(x,x_n)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(t-t_n)^2}{2\sigma^2}\right\}$$

$$= \sum_n k(x,x_n)\,\mathcal{N}(t|t_n,\sigma^2)$$

6.21 $C_N = \mathbb{K} + \beta^{-1}\mathbb{I}_N$

$$\mathbb{K} = \frac{1}{\alpha}\boldsymbol{\Phi}\boldsymbol{\Phi}^T \qquad \mathbf{k} = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_{N+1}) \\ \vdots \\ k(\mathbf{x}_N, \mathbf{x}_{N+1}) \end{pmatrix} = \frac{1}{\alpha}\boldsymbol{\Phi}\,\boldsymbol{\phi}(\mathbf{x}_{N+1})$$

(6.66)より.

$$m(\mathbf{x}_{N+1}) = \frac{1}{\alpha}\boldsymbol{\phi}(\mathbf{x}_{N+1})^T\boldsymbol{\Phi}^T\left[\frac{1}{\alpha}\boldsymbol{\Phi}\boldsymbol{\Phi}^T + \beta^{-1}\mathbb{I}_N\right]^{-1}\mathbf{t}$$

$$= \frac{1}{\alpha}\boldsymbol{\phi}(\mathbf{x}_{N+1})^T\alpha\beta\left[\beta\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \alpha\mathbb{I}_N\right]^{-1}\boldsymbol{\Phi}^T\mathbf{t}$$

$$= \beta\,\boldsymbol{\phi}(\mathbf{x}_{N+1})^T S_N\boldsymbol{\Phi}^T\mathbf{t}$$

6.25

$$\mathbf{a}_N^{\text{new}} = \mathbf{a}_N - (-W_N - C_N^{-1})^{-1}(\mathbf{t}_N - \boldsymbol{\sigma}_N - C_N^{-1}\mathbf{a}_N)$$

$$= \mathbf{a}_N + (W_N + C_N^{-1})^{-1}(\mathbf{t}_N - \boldsymbol{\sigma}_N - C_N^{-1}\mathbf{a}_N)$$

$$= (W_N + C_N^{-1})^{-1}\left\{(W_N + C_N^{-1})\mathbf{a}_N + \mathbf{t}_N - \boldsymbol{\sigma}_N - C_N^{-1}\mathbf{a}_N\right\}$$

$$= (W_N + C_N^{-1})^{-1}\left\{W_N\mathbf{a}_N + \mathbf{t}_N - \boldsymbol{\sigma}_N\right\}$$

$$= C_N C_N^{-1}(W_N + C_N^{-1})^{-1}(\mathbf{t}_N - \boldsymbol{\sigma}_N + W_N\mathbf{a}_N)$$

$$= C_N(C_N W_N + \mathbb{I})^{-1}(\mathbf{t}_N - \boldsymbol{\sigma}_N + W_N\mathbf{a}_N)$$

6.26 $p(a_{N+1}|\mathbf{t}_N) = \int p(a_{N+1}|\mathbf{a}_N)\,p(\mathbf{a}_N|\mathbf{t}_N)\,d\mathbf{a}_N$

$$= \int \mathcal{N}(a_{N+1}|\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{a}_N,\, c-\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k}) \cdot \mathcal{N}(\mathbf{a}_N|\mathbf{a}_N^*, \mathbf{H}^{-1})\,d\mathbf{a}_N$$

$$p(\mathbf{a}_N|\mathbf{t}_N) = \mathcal{N}(\mathbf{a}_N|\mathbf{a}_N^*, \mathbf{H}^{-1}).$$

$$p(a_{N+1}|\mathbf{a}_N) = \mathcal{N}(a_{N+1}|\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{a}_N,\, c-\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k})$$

$$p(a_{N+1}|\mathbf{t}_N) = \mathcal{N}\Big(a_{N+1}\Big|\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{a}_N^* = \mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{C}_N(\mathbf{t}_N-\boldsymbol{\sigma}_N),$$

$$c-\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{k}+\mathbf{k}^T\mathbf{C}_N^{-1}\mathbf{H}^{-1}(\mathbf{k}^T\mathbf{C}_N^{-1})^T$$

$$= c-\mathbf{k}^T(\mathbf{C}_N^{-1}-\mathbf{C}_N^{-1}\mathbf{H}^{-1}\mathbf{C}_N^{-1})\mathbf{k}$$

$$= c-\mathbf{k}^T(\mathbf{C}_N^{-1}-(\mathbf{C}_N\mathbf{W}_N\mathbf{C}_N+\mathbf{C}_N^{-1})^{-1})\mathbf{k}.\Big)$$ ☐