PRML勉強会(中嶋、6/21)

2020年6月14日　　14:30

1.9　(1.46) $\mathcal{N}(x|\mu,\sigma^2)=\dfrac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

のモードが $x=\mu$ であること

$$\frac{dN}{dx}=\frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\times\left[-\frac{1}{\sigma^2}(x-\mu)\right]$$ これが0となるのは $x=\mu$.

$$\frac{d^2N}{dx^2}=\frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\times\left(-\frac{1}{\sigma^2}(x-\mu)\right)^2$$

$$-\frac{1}{\sigma^2}\frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$x=\mu$ を代入すると $\dfrac{d^2N}{dx^2}<0$　∴ $x=\mu$ で N は最大

(1.52) $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\Sigma)=\dfrac{1}{(2\pi)^{D/2}}\dfrac{1}{|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$

$$\nabla_{\mathbf{x}}N=\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

$$\times\left[-\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$ $=0$ となるのは $\mathbf{x}=\boldsymbol{\mu}$ のとき.

ただし (C.19) $\dfrac{\partial}{\partial\mathbf{x}}(\mathbf{x}^T\mathbf{a})=\mathbf{a}$ を使用. また $\nabla e^{f(x)}=e^{f(x)}\nabla f(x)$ を使用.

ヘッセ行列について考える.

$\exp[\cdots]$ を x_i で微分して出てくる項をまず考える.

$$\begin{array}{cc} \left[-\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]\frac{\partial}{\partial x_1}\exp[\cdots] & \left[-\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]\frac{\partial}{\partial x_2}\exp[\cdots]\ \cdots \\ \text{1列目} & \text{2列目} \end{array}$$

$\mathbf{x}=\boldsymbol{\mu}$ を代入すると零行列.

一方, $-\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$ を x_i で微分した項は $\dfrac{\partial}{\partial x_i}\left[-\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$

$$=-\Sigma^{-1}\frac{\partial}{\partial x_i}(\mathbf{x}-\boldsymbol{\mu})$$

$$=-\Sigma^{-1}\begin{array}{cc}\begin{pmatrix}0\\0\\\vdots\\1\\\vdots\\0\end{pmatrix} & \begin{array}{c}\\\\\\\leftarrow i\text{行目}\\\\\\\end{array}\end{array}$$

従ってヘッセ行列全体としては

$$\begin{array}{cc}\left(\frac{\partial}{\partial x_1}\nabla N\right. & \left.\frac{\partial}{\partial x_2}\nabla N\ \cdots\right) \\ \text{1列目} & \text{2列目}\end{array}\propto-\left(\Sigma^{-1}\begin{pmatrix}1\\\vdots\\0\end{pmatrix}\ \Sigma^{-1}\begin{pmatrix}0\\\vdots\\0\end{pmatrix}\cdots\right)$$

$$=-\Sigma^{-1}$$

精度行列は 正定値. よって N は $\mathbf{x}=\boldsymbol{\mu}$ で最頻値をとる.

1.10　$E[x+z]=E[x]+E[z]$

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

を x と z が独立であるとき示せ.

x と z の周辺分布を $q(x), r(z)$ とする. 定義より $E[x]=\int xq(x)dx$, $E[z]=\int zr(z)dz$

ただし積分範囲は全実数.

$p(x,z)=q(x)r(z)$　(x と z は独立)

$$E[x+z]=\iint(x+z)p(x,z)dxdz$$
$$=\iint(x+z)q(x)r(z)dxdz$$
$$=\iint xq(x)r(z)dxdz+\iint zq(x)r(z)dxdz$$
$$=E[x]+E[y]$$

同様に $\mathrm{var}[x]=\int(x-E[x])^2q(x)dx$, $\mathrm{var}[z]=\int(z-E[z])^2r(z)dz$

$$\mathrm{var}[x+z]=\iint(x+z-E[x+z])^2p(x,z)dxdz$$
$$=\iint[(x-E[x])+(z-E[z])]^2q(x)r(z)dxdz$$
$$=\iint(x-E[x])^2q(x)r(z)dxdz+\iint(z-E[z])^2q(x)r(z)dxdz+\underset{\underset{0}{\shortparallel}}{\text{クロスターム}}$$

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

1.11　(1.54)から(1.55)と(1.56)を導出

(1.54) $\ln p(\mathbf{x}|\mu,\sigma^2)=-\dfrac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2-\dfrac{N}{2}\ln\sigma^2-\dfrac{N}{2}\ln(2\pi)$

μ で微分

$$\frac{\partial}{\partial\mu}\ln p(\cdots)=\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)\quad\cdots①$$

σ^2 で微分

$$\frac{\partial}{\partial\sigma^2}\ln p(\cdots)=\frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2-\frac{N}{2\sigma^2}\quad\cdots②$$

微分を0とおくと ① より

$$\sum_{n=1}^{N}(x_n-\mu)=\sum_{n=1}^{N}x_n-N\mu_{ML}=0\quad\therefore\mu_{ML}=\frac{1}{N}\sum_{n=1}^{N}x_n$$

② より

$$\frac{1}{2\sigma_{ML}^4}\sum_{n=1}^{N}(x_n-\mu_{ML})^2-\frac{N}{2\sigma_{ML}^2}=0$$

$$N\sigma_{ML}^2=\sum_{n=1}^{N}(x_n-\mu_{ML})^2\quad\therefore\sigma_{ML}^2=\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$

1.12　$E[x_nx_m]=\mu^2+I_{nm}\sigma^2$ を示せ.

ただし x_n, x_m は $\mathcal{N}(x|\mu,\sigma^2)$ から独立に($n\neq m$ のとき)サンプルされたデータ点で I_{nm} は単位行列の (n,m) 成分.

① $n=m$ のとき $E[x_nx_m]=E[x_n^2]=\mu^2+\sigma^2$ (1.50より)

② $n\neq m$ のとき

$$E[x_nx_m]=\iint x_nx_mp(x_n,x_m)dx_ndx_m$$
$$=\iint x_nx_m\mathcal{N}(x_n|\mu,\sigma^2)\mathcal{N}(x_m|\mu,\sigma^2)dx_ndx_m$$
$$=\int x_n\mathcal{N}(x_n|\mu,\sigma^2)dx_n\int x_m\mathcal{N}(x_m|\mu,\sigma^2)dx_m$$
$$=\mu^2$$

(1.57)の証明

$$E[\mu_{ML}]=E\left[\frac{1}{N}\sum_{n=1}^{N}x_n\right]=\frac{1}{N}\sum_{n=1}^{N}E[x_n]=\frac{1}{N}\sum_{n=1}^{N}\mu=\mu$$

$$E[\sigma_{ML}^2]=\frac{1}{N}\sum_{n=1}^{N}E[(x_n-\mu_{ML})^2]$$
$$=\frac{1}{N}\sum_{n=1}^{N}\left(E[x_n^2]-2E[x_n\mu_{ML}]+E[\mu_{ML}^2]\right)$$
$$=\frac{1}{N}\sum_{n=1}^{N}\left\{\mu^2+\sigma^2-2\mu^2-\frac{2}{N}\sigma^2+\frac{1}{N^2}\sum_{i,j}E[x_ix_j]\right\}$$
$$=\frac{1}{N}\sum_{n=1}^{N}\left\{\cancel{\mu^2}+\sigma^2-\cancel{2\mu^2}-\frac{2}{N}\sigma^2+\frac{1}{N^2}\left[N(\cancel{\mu^2}+\sigma^2)\cancel{+\partial N^2}-N\mu\right]\right\}$$
$$=\frac{1}{N}\times N\times\left(1-\frac{1}{N}\right)\sigma^2$$
$$=\frac{N-1}{N}\sigma^2$$