ML-HW2

1. Bayesian Inference

$$p(\Lambda|X) \propto \left(p(\Lambda)\prod_{n=1}^{N-1} p(x_n|\Lambda)\right)\cdot p(x_N|\Lambda)$$

$$= p(\Lambda)\cdot p(x_1|\Lambda)\cdot p(x_2|\Lambda)\cdots\cdots p(x_N|\Lambda)$$

$$= p(\Lambda)\cdot\prod_{n=1}^{N} p(x_n|\Lambda)$$

$$= p(\Lambda)\cdot\prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\Sigma)$$

$$\propto p(\Lambda)\cdot\Lambda^{\frac{N}{2}}\cdot e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu)}$$

$$= \mathcal{W}(\Lambda|W_0,\nu_0)\cdot\Lambda^{\frac{N}{2}}\cdot e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu)}$$

$$= \left(B\cdot|\Lambda|^{\nu_0-D-1/2}\cdot e^{-\frac{1}{2}Tr(W_0^{-1}\Lambda)}\right)\cdot\Lambda^{\frac{N}{2}}\cdot e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu)}$$

$$= B\cdot|\Lambda|^{\frac{\nu_0+N-D-1}{2}}\cdot e^{-\frac{1}{2}Tr(W_0^{-1}\Lambda)+\underbrace{-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu)}}$$

又 $\because \sum_{n=1}^{N}(x_n-\mu)^T\Lambda(x_n-\mu)$

$$= \sum_{n=1}^{N}[x_{n1}-\mu_1, x_{n2}-\mu_2,\ldots\ldots, x_{nD}-\mu_D]_{1\times D} * \Lambda_{D\times D} * \begin{bmatrix} x_{n1}-\mu_1 \\ x_{n2}-\mu_2 \\ \vdots \\ x_{nD}-\mu_D\end{bmatrix}_{D\times 1}$$, which D is the dimension of x-vector.

$$= \sum_{n=1}^{N}[(x_{n1}-\mu_1)\Lambda_{11},(x_{n2}-\mu_2)\Lambda_{22},\ldots\ldots,(x_{nD}-\mu_D)\Lambda_{DD}]_{1\times D} * \begin{bmatrix} \vdots \end{bmatrix}_{D\times 1}$$, suppose x_{ni}, x_{nj} is independent, $\forall i\neq j,\ i,j\in D$

$$= \sum_{n=1}^{N}\left((x_{n1}-\mu_1)(x_{n1}-\mu_1)\Lambda_{11}+(x_{n2}-\mu_2)(x_{n2}-\mu_2)\Lambda_{22}+\cdots+(x_{nD}-\mu_D)(x_{nD}-\mu_D)\Lambda_{DD}\right)$$

$$= \sum_{n=1}^{N} Tr(K\cdot\Lambda)$$, which $K = \begin{bmatrix} (x_{n1}-\mu_1)(x_{n1}-\mu_1) & 0 & \cdots & 0 \\ 0 & (x_{n2}-\mu_2)(x_{n2}-\mu_2) & & \\ \vdots & & \ddots & \end{bmatrix}_{D\times D}$

$$\therefore p(\Lambda|X) \propto B\cdot|\Lambda|^{\frac{(\nu_0+N)-D-1}{2}}\cdot e^{-\frac{1}{2}Tr(W_0^{-1}\Lambda)+-\frac{1}{2}\sum_{n=1}^{N}Tr(K\cdot\Lambda)}$$

$$= B\cdot|\Lambda|^{\frac{\nu_0+N-D-1}{2}}\cdot e^{-\frac{1}{2}Tr\left(W_0^{-1}\Lambda+\sum_{n=1}^{N}K\cdot\Lambda\right)}$$

$\because \Lambda$ is not relate to n #

$$= B \cdot |\Lambda|^{\frac{(\nu_0+N)-D-1}{2}} \cdot e^{-\frac{1}{2}\mathrm{Tr}\left(W_0^{-1}\Lambda + \mathbb{K}\Lambda\right)}$$, which $\mathbb{K} = \sum_{n=1}^{N} K$ #

$$= \mathcal{W}(\Lambda \mid W_\Lambda, \nu_\Lambda)$$, which

$$\begin{cases} W_\Lambda = W_0^{-1} + \mathbb{K} \\ \nu_\Lambda = \nu_0 + N \end{cases}$$