The 3rd quiz in Pattern Recognition and Machine Learning

(Answer)

Ex. 4.1

4.1 Solutions

4.1.1 Uncorrelated does not imply independent

We have

$$E[X] = 0 \tag{4.1}$$

$$\text{var}[X] = (1 - -1)^2/12 = 1/3 \tag{4.2}$$

$$E[Y] = E[X^2] = \text{var}[X] + (E[X])^2 = 1/3 + 0 \tag{4.3}$$

$$E[XY] = \int p(x)x^3dx = \frac{1}{2}\left[-(\int_{-1}^{0} x^3dx) + (\int_{0}^{1} x^3dx)\right] \tag{4.4}$$

$$= \frac{1}{2}\left[\frac{-1}{4}[x^4]_{-1}^{0} + \frac{1}{4}[x^4]_{0}^{1}\right] = \frac{1}{8}[-1+1] = 0 \tag{4.5}$$

where we have split the integral into two pieces, since x^3 changes sign in the interval $[-1, 1]$. Hence

$$\rho = \frac{\text{cov}[X,Y]}{\sigma_X\sigma_Y}\frac{E[XY] - E[X]E[Y]}{\sigma_X\sigma_Y} = \frac{0}{\sigma_X\sigma_Y} = 0 \tag{4.6}$$

4.1.5 Normalization constant for a multidimensional Gaussian

Ex. 4.1

Let $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, so

$$\mathbf{\Sigma}^{-1} = \mathbf{U}^{-T}\mathbf{\Lambda}^{-1}\mathbf{U}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U} = \sum_{i=1}^{p} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T \tag{4.30}$$

Hence

$$(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{p} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T\right)(\mathbf{x}-\boldsymbol{\mu}) \tag{4.31}$$

$$= \sum_{i=1}^{p} \frac{1}{\lambda_i}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{u}_i\mathbf{u}_i^T(\mathbf{x}-\boldsymbol{\mu}) = \sum_{i=1}^{p}\frac{y_i^2}{\lambda_i} \tag{4.32}$$

where $y_i \triangleq \mathbf{u}_i^T(\mathbf{x}-\boldsymbol{\mu})$. The $\mathbf{y}$ variables define a new coordinate system that is shifted (by $\boldsymbol{\mu}$) and rotated (by $\mathbf{U}$) with respect to the original x coordinates: $\mathbf{y} = \mathbf{U}(\mathbf{x}-\boldsymbol{\mu})$. Hence $\mathbf{x} = \mathbf{U}^T\mathbf{y} + \boldsymbol{\mu}$.

The Jacobian of this transformation, from $\mathbf{y}$ to $\mathbf{x}$, is a matrix with elements

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = U_{ji} \tag{4.33}$$

so $\mathbf{J} = \mathbf{U}^T$ and $|\mathbf{J}| = 1$.

So

$$\int \exp(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))d\mathbf{x} = \int\prod_i \exp(-\frac{1}{2}\sum_i \frac{y_i^2}{\lambda_i})dy_i|\mathbf{J}| \tag{4.34}$$

$$= \prod_i \sqrt{2\pi\lambda_i} = |2\pi\mathbf{\Sigma}| \tag{4.35}$$

Ex. 4.17 1 2

4.1.7 Conditioning a bivariate Gaussian

1. We have that $P(X_2|x_1) = \mathcal{N}(X_2|\mu_{2|1}, \Sigma_{2|1})$, where

$$\mu_{2|1} = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_1-\mu_1) = \mu_2 + \frac{\sigma_2\rho}{\sigma_1}(x_1-\mu_1)$$

$$\sigma_{2|1}^2 = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_2^2} = \sigma_2^2(1-\rho^2)$$

2. Now $\sigma_{12} = \rho$, so the above becomes

$$\mu_{2|1} = \mu_2 + \rho(x_1-\mu_1)$$

$$\sigma_{2|1}^2 = 1-\rho^2$$

Ex. 4.11

4.1.11 Derivation of the NIW posterior

Multiplying prior and likelihood we get

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathcal{D}) \propto |\boldsymbol{\Sigma}|^{-\frac{\nu_0+D+2+N}{2}} \mathrm{etr}\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\mathbf{S}_{\overline{x}} + \mathbf{S}_0 + \right.\right. \tag{4.65}$$

$$\left.\left. +N(\boldsymbol{\mu} - \overline{\mathbf{x}})(\boldsymbol{\mu} - \overline{\mathbf{x}})^T + \kappa_0(\boldsymbol{\mu} - \mathbf{m}_0)(\boldsymbol{\mu} - \mathbf{m}_0)^T\right)\right] \tag{4.66}$$

where $\mathrm{etr}(\mathbf{M}) = \exp(\mathrm{tr}(\mathbf{M}))$. Using the hint this becomes

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathcal{D}) \propto |\boldsymbol{\Sigma}|^{-\frac{\nu_N+D+2}{2}} \mathrm{etr}\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\mathbf{S}_{\overline{x}} + \mathbf{S}_0 + \frac{\kappa_0 N}{\kappa_N}(\overline{\mathbf{x}} - \mathbf{m}_0)(\overline{\mathbf{x}} - \mathbf{m}_0)^T\right)\right] \tag{4.67}$$

$$\times \exp\left[-\frac{\kappa_N}{2}(\boldsymbol{\mu} - \mathbf{m}_N)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{m}_N)\right] \tag{4.68}$$

For the NWI case,

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\mathcal{D}) = \mathrm{NWI}(\boldsymbol{\mu}, \boldsymbol{\Lambda}|\mathbf{m}_n, \kappa_n, \nu_n, \mathbf{W}_n) \tag{4.69}$$

$$\mathbf{m}_n = \frac{\kappa_0\mathbf{m}_0 + N\overline{\mathbf{x}}}{\kappa_n} \tag{4.70}$$

$$\kappa_n = \kappa_0 + N \tag{4.71}$$

$$\nu_n = \nu_0 + N \tag{4.72}$$

$$\mathbf{W}_n^{-1} = \mathbf{W}_0^{-1} + \mathbf{S}_{\overline{x}} + \frac{\kappa_0 N}{\kappa_0 + N}(\overline{\mathbf{x}} - \mathbf{m}_0)(\overline{\mathbf{x}} - \mathbf{m}_0)^T \tag{4.73}$$