26 June 2023

The 3rd quiz in Pattern Recognition and Machine Learning

Due 3 July 2023

Exercises

Exercise 4.1 Uncorrelated does not imply independent

Let $X \sim U(-1, 1)$ and $Y = X^2$. Clearly Y is dependent on X (in fact, Y is uniquely determined by X). However, show that $\rho(X, Y) = 0$. Hint: if $X \sim U(a, b)$ then $E[X] = (a + b)/2$ and $\text{var}[X] = (b - a)^2/12$.

Exercise 4.5 Normalization constant for a multidimensional Gaussian

Prove that the normalization constant for a d-dimensional Gaussian is given by

$$(2\pi)^{d/2}|\mathbf{\Sigma}|^{\frac{1}{2}} = \int \exp(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))d\mathbf{x} \tag{4.265}$$

Hint: diagonalize $\mathbf{\Sigma}$ and use the fact that $|\mathbf{\Sigma}| = \prod_i \lambda_i$ to write the joint pdf as a product of d one-dimensional Gaussians in a transformed coordinate system. (You will need the change of variables formula.) Finally, use the normalization constant for univariate Gaussians.

Exercise 4.7 Conditioning a bivariate Gaussian

Consider a bivariate Gaussian distribution $p(x_1, x_2) = \mathcal{N}(x|\mu, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \sigma_1\sigma_2 \begin{pmatrix} \frac{\sigma_1}{\sigma_2} & \rho \\ \rho & \frac{\sigma_2}{\sigma_1} \end{pmatrix} \tag{4.269}$$

where the correlation coefficient is given by

$$\rho \triangleq \frac{\sigma_{12}}{\sigma_1\sigma_2} \tag{4.270}$$

a. What is $P(X_2|x_1)$? Simplify your answer by expressing it in terms of ρ, σ_2, σ_1, μ_1,μ_2 and x_1.

b. Assume $\sigma_1 = \sigma_2 = 1$. What is $P(X_2|x_1)$ now?

Exercise 4.11 Derivation of the NIW posterior

Derive Equation 4.209. Hint: one can show that

$$N(\overline{\mathbf{x}} - \boldsymbol{\mu})(\overline{\mathbf{x}} - \boldsymbol{\mu})^T + \kappa_0(\boldsymbol{\mu} - \mathbf{m}_0)(\boldsymbol{\mu} - \mathbf{m}_0)^T \tag{4.271}$$

$$= \kappa_N(\boldsymbol{\mu} - \mathbf{m}_N)(\boldsymbol{\mu} - \mathbf{m}_N)^T + \frac{\kappa_0 N}{\kappa_N}(\overline{\mathbf{x}} - \mathbf{m}_0)(\overline{\mathbf{x}} - \mathbf{m}_0)^T \tag{4.272}$$

This is a matrix generalization of an operation called **completing the square**.[5]

Derive the corresponding result for the normal-Wishart model.

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathcal{D}) = \mathrm{NIW}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{m}_N, \kappa_N, \nu_N, \mathbf{S}_N) \tag{4.209}$$

$$\mathbf{m}_N = \frac{\kappa_0 \mathbf{m}_0 + N\overline{\mathbf{x}}}{\kappa_N} = \frac{\kappa_0}{\kappa_0 + N}\mathbf{m}_0 + \frac{N}{\kappa_0 + N}\overline{\mathbf{x}} \tag{4.210}$$

$$\kappa_N = \kappa_0 + N \tag{4.211}$$

$$\nu_N = \nu_0 + N \tag{4.212}$$

$$\mathbf{S}_N = \mathbf{S}_0 + \mathbf{S}_{\overline{x}} + \frac{\kappa_0 N}{\kappa_0 + N}(\overline{\mathbf{x}} - \mathbf{m}_0)(\overline{\mathbf{x}} - \mathbf{m}_0)^T \tag{4.213}$$

$$= \mathbf{S}_0 + \mathbf{S} + \kappa_0 \mathbf{m}_0 \mathbf{m}_0^T - \kappa_N \mathbf{m}_N \mathbf{m}_N^T \tag{4.214}$$

where we have defined $\mathbf{S} \triangleq \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T$ as the uncentered sum-of-squares matrix (this is easier to update incrementally than the centered version).