Machine Learning Worksheet 4

Multivariate Normal

1 Basic Gaussian

Problem 1: We decomposed the covariance matrix $\boldsymbol{\Sigma}$ into $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ with the orthonormal matrix $\boldsymbol{U}$ and the diagonal matrix of eigenvalues $\boldsymbol{\Lambda}$. Show that:

$$\boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T = \sum_{i=1}^{d} \frac{1}{\lambda_i} \boldsymbol{u}_i \boldsymbol{u}_i^T$$

We know that:

Index notation for outer product:

$$(\boldsymbol{AB})_{ij} = \sum_k a_{ik} b_{kj}$$

Index notation for matrix product:

$$(\boldsymbol{u}\boldsymbol{v}^T)_{ij} = u_i u_j$$

Indexing of column vectors:

$$(\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_k) = U$$

On the LHS we have:

$$(\boldsymbol{U}\boldsymbol{\Lambda})_{ij} = \sum_k u_{ik} \frac{1}{\lambda_{kj}} = u_{ij} \frac{1}{\lambda_j}$$

$$\left((\boldsymbol{U}\boldsymbol{\Lambda})\,\boldsymbol{U}^T\right)_{ij} = \sum_k u_{ik} \frac{1}{\lambda_k} u_{jk} = \sum_k \frac{1}{\lambda_k} u_{ik} u_{jk}$$

On the RHS we have:

$$\left(\sum_k \frac{1}{\lambda_k} \boldsymbol{u}_k \boldsymbol{u}_k^T\right)_{ij} = \sum_k \frac{1}{\lambda_k} \left(\boldsymbol{u}_k \boldsymbol{u}_k^T\right)_{ij} = \sum_k \frac{1}{\lambda_k} \boldsymbol{u}_{ik} \boldsymbol{u}_{jk}$$

Problem 2: Let $\boldsymbol{L}$ be an invertible n by n matrix over $\mathbb{R}$. Show that the linear transformation $Y = \boldsymbol{L}X$ of a Gaussian is again a Gaussian. Use the moment parametrisation and the change of variable theorem to get there.

Moment parametrisation of a Gaussian:

$$\propto e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

We apply the linear transformation $y = \boldsymbol{L}x$. $\boldsymbol{x}(\boldsymbol{y})$ is therefore $x = \boldsymbol{L}^{-1}y$

Together with the change of variable theorem we get:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}(\boldsymbol{y}))\left|\frac{\boldsymbol{dx}}{\boldsymbol{dy}}\right| = f(\boldsymbol{L}^{-1})\,|\boldsymbol{L}| \propto e^{-\frac{1}{2}(\boldsymbol{L}^{-1}y-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{L}^{-1}y-\boldsymbol{\mu})}$$
$$= e^{-\frac{1}{2}(\boldsymbol{L}^{-1}(y-\boldsymbol{L\mu}))^T\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{-1}(y-\boldsymbol{L\mu})} = e^{-\frac{1}{2}(y-\boldsymbol{L\mu})^T\boldsymbol{L}^{-T}\boldsymbol{\Sigma}^{-1}\boldsymbol{L}^{-1}(y-\boldsymbol{L\mu})}$$
$$= e^{-\frac{1}{2}(y-\boldsymbol{L\mu})^T(\boldsymbol{L\Sigma L}^T)^{-1}(y-\boldsymbol{L\mu})}$$

2 Conditioned Gaussian

You got a dataset with pairs $\boldsymbol{x}$ and $\boldsymbol{y}$. You computed the covariance of the concatenated datapoints $\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}$. The covariance matrix looks like this: $\begin{pmatrix}\boldsymbol{\Sigma}_{X,X} & \boldsymbol{\Sigma}_{X,Y}\\ \boldsymbol{\Sigma}_{Y,X} & \boldsymbol{\Sigma}_{Y,Y}\end{pmatrix}$. For simplicity we assume that $\boldsymbol{\mu}_X$ and $\boldsymbol{\mu}_Y$ are $\mathbf{0}$.

Problem 3: The underlying process of the probability distribution $p(\boldsymbol{y} \mid \boldsymbol{x})$ can be expressed as $Y = \boldsymbol{T}X + Z$ where $Z \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{Y|X})$. What is the value of the matrix $\boldsymbol{T}$?

From Bishop we know that

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z)$$
$$\boldsymbol{\mu}_Z = \begin{pmatrix}\boldsymbol{\mu}_X\\ \boldsymbol{A\mu}_X + \boldsymbol{b}\end{pmatrix}$$
$$\boldsymbol{\Sigma}_Z = \begin{pmatrix}\boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_X\boldsymbol{A}^T\\ \boldsymbol{A\Sigma}_X & \boldsymbol{\Sigma}_{Y|X} + \boldsymbol{A\Sigma}_X\boldsymbol{A}^T\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{X,X} & \boldsymbol{\Sigma}_{X,Y}\\ \boldsymbol{\Sigma}_{Y,X} & \boldsymbol{\Sigma}_{Y,Y}\end{pmatrix}$$

In our transformation T corresponds to A, so we need to solve:

$$\boldsymbol{\Sigma}_X\boldsymbol{T}^T = \boldsymbol{\Sigma}_{X,Y}$$
$$\boldsymbol{T} = (\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{X,Y})^T$$

Problem 4: What is the covariance of Z?

From the last question we know that:

$$\boldsymbol{\Sigma}_{Y|X} + \boldsymbol{T}\boldsymbol{\Sigma}_X\boldsymbol{T}^T = \boldsymbol{\Sigma}_{Y,Y}$$

$$\boldsymbol{\Sigma}_{Y|X} = \boldsymbol{\Sigma}_{Y,Y} - \boldsymbol{T}\boldsymbol{\Sigma}_X\boldsymbol{T}^T$$