Tips for task 2

Exercise 6

A vector space is a space, where vectors can be added and multiplied with a scalar (where the scalar is usually an element of a field). Multiplication with a scalar does not imply that vectors can be multiplied. Think about the valid operations in a vector space. Especially:

Think if aa^T is the same as $a^T a$ and how it relates to matrices.

Exercise 7

a)

We want to decompose an arbitrary matrix $\mathbf{A}$ into:

$$A = W_S + W_a$$

where W_S is a symmetric matrix and W_a is skew symmetric.

We basically have the constraints:

$$\begin{aligned} w_{ij} &= w_{ij}^S + w_{ij}^A \\ w_{ji} &= w_{ji}^S + w_{ji}^A \end{aligned}.$$

Can these equations be solved while still retaining the symmetric properties of W?

c)

The is a great chance to use induction! Start with a simple example $D = 1$ and see how it generalizes to the general case. Especially: How does $\frac{D(D+1)}{2}$ relate to the Gaussian summation formula (Gauss sum).

Exercise 8

We have several ways to solve this exercise.

Way 1: Start off with Bayes' theorem:

$$
\begin{aligned}
& p(\mathbf{x}_a|\mathbf{x}_b) &&= \frac{p(\mathbf{x}_b|\mathbf{x}_a)\cdot p(\mathbf{x}_a)}{p(\mathbf{x}_b)} \\
\Leftrightarrow \quad & p(\mathbf{x}_a) &&= \frac{p(\mathbf{x}_a|\mathbf{x}_b)p(\mathbf{x}_b)}{p(\mathbf{x}_b|\mathbf{x}_a)} \\
& p(\mathbf{x}_a) &&= \frac{\underbrace{p(\mathbf{x}_a,\mathbf{x}_b)}_{joint\ pdf}}{\underbrace{p(\mathbf{x}_b|\mathbf{x}_a)}_{conditional\ pdf}}.
\end{aligned}
$$

Since we know the joint pdf $p(\mathbf{x}_a,\mathbf{x}_b)$ and the conditional pdf $p(\mathbf{x}_b|\mathbf{x}_a)$ (see slides or prove it for yourself), we can just calculate this out (Do not forget to treat the $\mathbf{x}_b$ as constants and not variables!)

Way 2:

We can also regain the original pdf by marginalization:

$$p(\mathbf{x}_a) = \int_{-\infty}^{\infty} (2\pi)^{-\frac{D}{2}} \left|\mathbf{\Sigma}^{-\frac{1}{2}}\right| \cdot e^{-\frac{1}{2}[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]}\, dx_b.$$

Inside the integral, we have to treat the variables $\mathbf{x}_a$ as constants and factor them out. The result should be something like:

$$\int_{-\infty}^{\infty}(2\pi)^{-\frac{D}{2}} \left|\mathbf{\Sigma}^{-\frac{1}{2}}\right| \cdot e^{-\frac{1}{2}[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]}\, dx_b = N(x_b|\boldsymbol{\mu}_a,\Sigma_{aa}) \underbrace{\int_{-\infty}^{\infty} k(x_b)\ dx_b}_{should\ integrate\ to\ one}.$$

This could be more involved than way 1, but is definitely possible.

Exercise 9

When we see a maximum likelihood task, we repeat these menial steps:

- Find Likelihood function $L(\boldsymbol{\theta})$
- Take the derivative $\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and solve $\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$.
- Check if the solution is a (global) maxima.

Example: Maximum likelihood estimator for variance in this task:

$$\prod_{n=1}^{N} p(x_n;\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2} = L(\sigma)$$

If we took the derivative of the above term, we would have to use the product rule. This can become unhandy in many situations, because each application of the product rule increases the number of terms by a factor of two.

Instead we take the logarithm. Because $L(\sigma)$ is always positive, it is well defined. Additionally (and more importantly), this will transform our product into a sum without changing the location of the wanted maxima.

$$\ln L(\sigma) = -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2.$$

Now we solve for zero. Do not forget the special rules for derivatives of logarithms!

$$\frac{\partial L(\sigma)}{\partial \sigma} = 0$$

$$\Leftrightarrow -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i - \mu)^2 = 0.$$

Solving for σ leads to:

$$\frac{1}{n}\sum_{i=1}^{N}(x_i - \mu)^2 = \sigma^2.$$

So the maximum likelihood estimate of the variance is just the sample variance. But in this case it is biased. See for yourself what that means (e.g. google Bessels correction). Then do the same for the mean.

Exercise 10

We have

$$\mathbf{y} = \mathbf{Ax} + \mathbf{b}.$$

As usual, we can solve this exercise in several ways. I use the change of variable formula:

If $y = g(\mathbf{x})$

Then:

$$f_Y(\mathbf{y}) = f_X\left(g^{-1}(\mathbf{y})\right) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|.$$

An explanation of the multiple ways to solve this can be found online (google: change of variable) or in my script “Stochastik für Informatiker” (found in Github, username “CowFreedom” or on Amazon (but Github is free!)) under the section “Transformation von Zufallsvariablen”.

In this case we have:

$$\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{b}) = g^{-1}(y).$$

This can now be plugged into the pdf f_X of $\mathbf{x}$:

$$f_y(y) = (2\pi)^{-\frac{D}{2}} \cdot \left|\boldsymbol{\Sigma}^{-\frac{1}{2}}\right| \cdot e^{-\frac{1}{2}\left[\left(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{b})-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{b})-\boldsymbol{\mu}\right)\right]} \cdot \left|\mathbf{A}^{-1}\right|$$

Do not forget, that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ for invertible matrices and

$\left|\mathbf{A}^{-1}\right| = \left|\left(\mathbf{A}^{1/2}\mathbf{A}^{1/2}\right)^{-1}\right|$ and $|\mathbf{A}| = |\mathbf{A}^T|$.

Ask yourself of the limits of the procedure. What if $\mathbf{A}$ is not invertible?