Probabilistic models for neural data
Session 2: Linear models

To do before this session:

- **Complete exercise 1 (see session 1 notes)**
- **Read up on this session's statistical concepts**
- **Complete pre-session quiz**

In this session we will first discuss the results of the first exercise. After that, we will focus on linear models, the use of priors for regularization, and on Bayesian model comparison. In Bayesian statistics, linear models and multivariate Gaussian distributions are tightly linked, as a linear transformation of a Gaussian random variable remains a Gaussian. Furthermore, a Gaussian prior on the mean of a Gaussian likelihood makes the posterior a Gaussian, leading to tractable Bayesian inference (an instance of a powerful statistical concept known as *conjugacy*). This makes Gaussians the single most important probability distribution to model continuous quantities. In the next session, we will revisit these concepts in Park & Pillow (2011), who apply them to estimating receptive fields from neural data.

Statistical concepts: multivariate Gaussian distributions and linear models

This week's statistical concepts are more involved, and require a good understanding of linear algebra. Given the importance of Gaussian distributions and linear models in general, make sure to spend sufficient time on trying to understand these concepts. There is a lot to read in preparation for this session, but you might be able to go through some parts quickly if you have some prior knowledge of linear models.

The Gaussian distribution (PRML 2.0, 2.3.0-2.3.7)
Make sure to understand

- The Gaussian's probability density function, and the meaning of iso-probability (hyper-)ellipsoids
- Completing the square to find the Gaussian parameters; the use of this technique comes up frequently when performing Bayesian inference with Gaussians
- Partitioning of Gaussians by marginalization; this is an essential property that makes Gaussian Processes possible (used in later sessions)
- Bayes rule as a special case of the conditional distribution
- Maximum likelihood estimates and Bayesian inference
- (Advanced) Conjugate priors for precision parameter and matrix
- (Advanced) Student's t distribution as a Gaussian with uncertain precision parameter

To save time, there is no need to go through the detailed derivations of the relevant sections. In general, it is good practice to re-derive the results at least once for yourself, to gain a better intuitive understanding of the underlying concepts. The advanced topics are not essential to understand the next session's paper, but it is good to have heard of them for later sessions.

Non-Bayesian and Bayesian linear models (PRML 3.0-3.1, 3.3)
Make sure to understand

- Maximum likelihood and least squares estimates, and their relationship
- The different impact of quadratic (Gaussian prior) and Lasso (Laplace prior) regularizers
- The impact of prior choices on Bayesian posteriors

It is sufficient to only skim the geometric interpretation and sequential learning sections. You might want to revisit them at a later stage to gain a deeper understanding of linear models. You can furthermore skim the section on the equivalent kernel (PRML 3.3.3), but you should revisit it once we discuss Gaussian processes.

Bayesian model comparison (PRML 3.4)
This section provides some insight into why Bayesian model comparison works, and how it penalizes more complex models that can potentially model a wider range of data. Note that Bayesian model comparison assumes that the used model is correct. A safer approach to model comparison are posterior predictive checks (e.g., cross validation) that do not make this assumption, but might require more data.

The evidence approximation (PRML 3.5)
The paper we will discuss in the next session employs the evidence approximation, also known as empirical Bayes. Feel free to skip the detailed derivations of this section, but you should understand why the evidence approximation was used, what assumptions it makes, and when/how these assumptions are violated.

Further reading / references
PRML Appendix B: an immensely useful summary of the properties of Gaussians and other distributions
Cheat-sheet on multivariate Gaussian distribution and linear models (Kevin Murphy): https://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/reading/gauss.pdf
Cheat-sheet on conjugate prior distributions for Gaussians and associated posterior (Kevin Murphy): https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf
Chapter 28 of MacKay's book (see Session 1 notes) has a good discussion on Bayesian model comparison, upon which PRML 3.4 is based.