Computational Neuroscience

Lecturer: Joao Barbosa
joao.barbosa@inserm.fr
Institute for Neuromodulation, Paris

Teaching Assistant: Pierre Houzelstein
pierre.houzelstein@gmail.com
École Normale Supérieure
29 rue d'Ulm, 2nd floor

Exercise Sheet 2 — 9 April 2025

Please submit your solution in the next class with Joao (5 May 2025)

1. **Covariance and Correlation**. Assume that we have recorded two neurons in the two-alternative-forced choice task discussed in class. We denote the firing rate of neuron 1 in trial i as $r_{1,i}$ and the firing rate of neuron 2 as $r_{2,i}$. We furthermore denote the averaged firing rate of neuron 1 as $\bar{r}_1$ and of neuron 2 as $\bar{r}_2$. Let us "center" the data by defining two data vectors

$$\mathbf{x} = \begin{pmatrix} r_{1,1} - \bar{r}_1 \\ r_{1,2} - \bar{r}_1 \\ r_{1,3} - \bar{r}_1 \\ \vdots \\ r_{1,N} - \bar{r}_1 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} r_{2,1} - \bar{r}_2 \\ r_{2,2} - \bar{r}_2 \\ r_{2,3} - \bar{r}_2 \\ \vdots \\ r_{2,N} - \bar{r}_2 \end{pmatrix}.$$

   (a) Show that the variance of the firing rates of the first neuron is

$$\mathrm{Var}(r_1) = \frac{1}{N-1}\|\mathbf{x}\|^2.$$

   (b) Compute the cosine of the angle between $\mathbf{x}$ and $\mathbf{y}$. What do you get?

   (c) What are the maximum and minimum values that the correlation coefficient between r_1 and r_2 can take? Why?

   (d) What do you think the term "centered" refers to?

2. **Ridge Regression – Closed-Form Solution**

   In class we discuss how to quantify overfitting by using cross-validation. In order to avoid overfitting, you should penalize complex models. This is called regularization. Here's a linear regression model with L2 regularization (ridge regression):

$$E(w) = \|y - Xw\|^2 + \lambda\|w\|^2$$

   where:

   - $X \in \mathbb{R}^{n \times d}$ is your data.
   - $y \in \mathbb{R}^n$ is the target vector.

- $w \in \mathbb{R}^d$ is the weight vector (i.e. the parameters of the model you are trying to fit).
- $\lambda \in \mathbb{R}^+$ is the regularization coefficient

(a) Why is the extra term $\lambda\|w\|^2$ penalizing complex models?

(b) Start by assuming $\lambda = 0$ (i.e. regression without regularization) and solve for the optimal weight vector w. Hint: as we did for encoding models, start by expanding the error function $E(w)$ and find its minimum. The solution should make use of the pseudo-inverse.

(c) Now make $\lambda > 0$ and solve using the same approach.

(d) If you found the right solution, you will add λI to $X^T X$. What problem does this solve? Hint: think about the rank of λI.

(e) What is the optimal value for λ?

3. **Bayes' theorem**. The theorem of Bayes summarizes all the knowledge we have about about the stimulus by observing the responses of a set of neurons, *independently* of the specific decoding rule. To get a better intuition about this theorem, we will look at the motion discrimination task again and compute the probability that the stimulus moved to the left ($\leftarrow$) or right ($\rightarrow$). For a stimulus $s = \{\leftarrow, \rightarrow\}$, and a firing rate response r of a single neuron, Bayes' theorem reads

$$p(s|r) = \frac{p(r|s)p(s)}{p(r)} \quad .$$

Here, $p(r|s)$ is the probability that the firing rate is r if the stimulus was s. The respective distribution can be measured and we assume that it follows a Gaussian probability density with mean μ_s and standard deviation σ,

$$p(r|s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r-\mu_s)^2}{2\sigma^2}\right)$$

The relative frequency with which the stimuli (leftward or rightward motion, $\leftarrow$ or $\rightarrow$) appear is denoted by $p(s)$, often called the *prior probability* or, for short, the *prior*. The distribution $p(r)$ denotes the probability of observing a response r, independent of any knowledge about the stimulus.

(a) How can you calculate $p(r)$? What shape does it have?

(b) The distribution $p(s|r)$ is often called the *posterior* probability or, for short, the *posterior*. Calculate the posterior for $s =\leftarrow$ and sketch it as a function of r, assuming a prior $p(\leftarrow) = p(\rightarrow) = 1/2$. Draw the posterior $p(\rightarrow|r)$ into the same plot.

(c) What happens if you change the prior? Investigate how the posterior changes if $p(\leftarrow)$ becomes much larger than $p(\rightarrow)$ and vice versa. Make a sketch similar to (b).

(d) Let us assume that we decide for leftward motion whenever $r > \frac{1}{2}(\mu_{\leftarrow} + \mu_{\rightarrow})$. Interpret this decision rule in the plots above. How well does this rule do depending on the prior? What do you lose when you move from the full posterior to a simple decision (decoding) rule?

4. **Linear discriminant analysis (advanced):** Let us redo the calculations for the case of N neurons. If we denote by $\mathbf{r}$ the vector of firing rates, Bayes' theorem reads:

$$p(s|\mathbf{r}) = \frac{p(\mathbf{r}|s)p(s)}{p(\mathbf{r})} \quad .$$

We assume that the distribution of firing rates again follows a Gaussian so that

$$p(\mathbf{r}|s) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_s)^T \mathbf{C}^{-1}(\mathbf{r}-\boldsymbol{\mu}_s)\right)$$

where $\boldsymbol{\mu}_s$ denotes the mean of the density for stimulus $s = \{\leftarrow, \rightarrow\}$, and $\mathbf{C}$ is the covariance matrix, assumed identical for both stimuli.

(a) Compute the log-likelihood ratio

$$l(\mathbf{r}) = \log \frac{p(\mathbf{r}\,|\!\leftarrow)}{p(\mathbf{r}\,|\!\rightarrow)} \quad .$$

(b) Assume that $l(\mathbf{r}) = 0$ is the decision boundary, so that any firing rate vector $\mathbf{r}$ giving a log-likelihood ratio larger than zero is classified as coming from the stimulus $\leftarrow$. Compute a formula for the decision boundary. What shape does this boundary have?

(c) Assume that $p(\leftarrow) = p(\rightarrow) = 1/2$. Assume we are analyzing two neurons with uncorrelated activities, so that the covariance matrix is

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Sketch the decision boundary for this case.