Data Analysis and Model Classification
Miniproject 3: Description and Assignmenent

Ruslan Aydarkhanov Bastien Orset Julien Rechenmann
Ricardo Chavarriaga José del R. Millán

November 20, 2017

Objective

In the previous miniproject, we have developed some *supervised learning* methods for feature selection and classification, where data samples were labeled as belonging to one of some known classes (e.g. malignant/benign or correct/erroneous samples). However, not all the problems we want to solve are discrete.

In this miniproject we will familiarize with regression (supervised learning) for the purpose of decoding continuous arm movements of a monkey from neural activities recorded from its brain.

Dataset description

The dataset for the third mini-project is adapted from an invasive brain machine interface (BMI) experiment (Fig. 1). In this experiment, a monkey is moving a pole with its arm while a number of electrodes implanted in the monkey's brain (multi-unit recording) are used to obtain neural firing rates corresponding to the arm movements. The neural activity is processed and decoded to predict the arm movement trajectories. The trajectories decoded from the neuronal data are then used to control an external robotic arm. So the pole is only used to correlate firing rates with arm trajectories, while the robot arm is controlled only by the brain via the developed decoder.

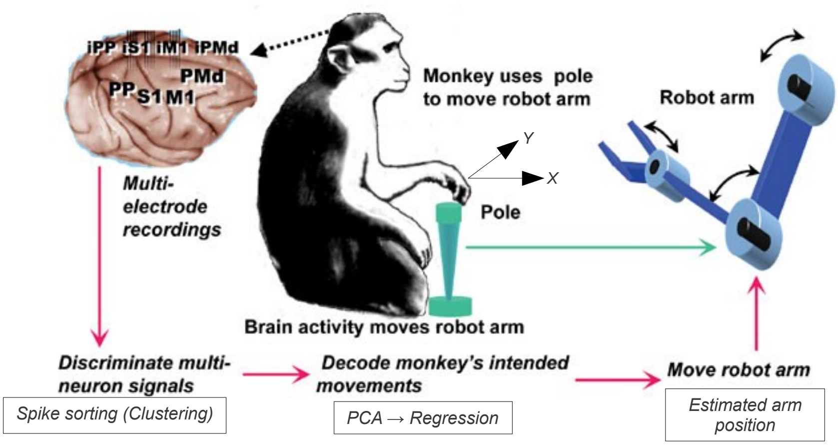

Figure 1: Experimental setup. Steps for data analysis are also shown.

The dataset *Data.mat* contains already the computed firing rates of all identified neurons. At time t, the dataset consists of the firing rates of 48 neurons recorded in a 1-second long window with a sampling frequency of 20 Hz (= number of spikes in 50 ms bins). Therefore, for each time t, the variable $\mathbf{Data}_t$ denotes the activity of 48 neurons with 20 samples ($20 \times 50\text{ms} = 1\text{s}$), resulting in $48 \times 20 = 960$ features for each time sample $\mathbf{Data}_t$. See the table 1 for a graphical representation.

Time in ms	**Data:** Spike rate of neurons $N^1 \dots N^{48}$										**PosX**	**PosY**
$t_0 = 0$	$N^1_{t_0-950\text{ms}}$	$N^2_{t_0-950\text{ms}}$	$\dots$	$N^{48}_{t_0-950\text{ms}}$	$N^1_{t_0-900\text{ms}}$	$\dots$	$N^{48}_{t_0-900\text{ms}}$	$N^1_{t_0-850\text{ms}}$	$\dots$	$N^{48}_{t_0}$	x_0	y_0
$t_1 = 50$	$N^1_{t_1-950\text{ms}}$	$N^2_{t_1-950\text{ms}}$	$\dots$	$N^{48}_{t_1-950\text{ms}}$	$N^1_{t_1-900\text{ms}}$	$\dots$	$N^{48}_{t_1-900\text{ms}}$	$N^1_{t_1-850\text{ms}}$	$\dots$	$N^{48}_{t_1}$	x_1	y_1
$t_2 = 100$	$N^1_{t_2-950\text{ms}}$	$N^2_{t_2-950\text{ms}}$	$\dots$	$N^{48}_{t_2-950\text{ms}}$	$N^1_{t_2-900\text{ms}}$	$\dots$	$N^{48}_{t_2-900\text{ms}}$	$N^1_{t_2-850\text{ms}}$	$\dots$	$N^{48}_{t_2}$	x_2	y_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$

Table 1: Arrangement of data variables in *Data.mat*. Note that the red cells contain the same values, since every time step shifts by a bin of 50 ms. This holds for all 48 neurons.

Apart from the firing rates, the arm movement trajectories $PosX_t$ and $PosY_t$ are also recorded during the experiment. The data for arm trajectories **PosX** and **PosY** is used as ground truth to design a regressor model where the input is neuronal activity. In other words, the firing rates $\mathbf{Data}_t$ will be compared with the actual hand trajectories $PosX_t$ and $PosY_t$ to predict hand movement trajectories from brain signals directly. The brain decoder can then be used to control the robotic arm, bypassing any muscular activity.

The data can be loaded into MATLAB from the file *Data.mat*. The loaded variables data are:
Data: Time samples × Features (48 neurons × 20 50-ms bins) of neuron spike rates
PosX: Cartesian coordinate X of monkey's wrist = pole
PosY: Cartesian coordinate Y of monkey's wrist = pole

Main Challenges (Checklist)

- Regress the arm movement trajectory to the neuronal firing rate. Your regression framework can use any of the methods seen in the guidesheets (e.g. PCA, regression, regularized regression). Make sure you justify your approach.
- Choose your final model and estimate its the performance. Put the exact model description and the obtained performance (which metric?) in the report!

The Report

At the end of the second mini-project you are requested to hand in a report, motivating the choices you made and discussing the results you obtained from your implementation. This report should cover **all aspects** of your analysis. You should demonstrate the understanding of how the robotic arm can be controlled with neural signals. We expect an integral consistent text rather than unrelated parts reflecting the guidesheets. The guidesheets are there to help you familiarize with different aspects of the data analysis. Try to combine the knowledge you acquired during the course. Please stick to the following structure:

Introduction In a few sentences (1-2 paragraphs), state the research question and the goal of the report. Give a **brief** description of the methods you have used (max. one page).

Methods Describe step by step and explain in detail what you did exactly, including

- If you have any hyper-parameters, how do you choose them (e.g. λ in Lasso regression)? (There are various ways of doing. If you choose to split your data into training set, validation set and testing set, there is no need to do cross-validation. The data is big enough to do a single partitioning.)
- Make sure it is always clear on which set (training set, testing set, ...) you perform the various steps (e.g. optimization, evaluation, ...). Use the correct nomenclature.

- How do you assess the performance of your model?
- The **final model**. Clearly state how and why you picked the methods you used to obtain the **one** final model.

Results This section is uniquely reserved for the results you obtained. Of course, it is not necessary to show the outcome of every single thing you have tried for this mini-project. So choose well what data to show and what not to show. You should be creative and pragmatic about how to display your data.

Discussion Use this chapter to briefly sum up your results and discuss about following points:

- What are the particularities of your dataset?
- Which methods do you recommend and why?
- Anything especially difficult or clever you did, and why?
- Any points where your methods did not suffice, and you would be in need of more advanced concepts.
- Choose your single **final** pipeline of the data processing. Report **all** the hyperparameters you have chosen. How would you implement the control of robotic arm based on the electrical brain signals? Don't forget to state your estimation of the performance of your model un unseen data?

Note: Your report **must not exceed 6 pages** (excluding the titlepage)! Every additional page will not be considered for grading.

Hint: Use the feedback you got on your first two reports to make this one even better! :)

Submission

The **deadline** for the report submission is **December 4, 2017 at 23H59**. Make sure your report is in .pdf format and name it according to following convention: `Miniproject3_Group`⟨*groupnumber*⟩`.pdf`, e.g. *Miniproject3_Group99.pdf*. Please also attach any MATLAB code used to obtain the results in a .zip file. The code will not be graded and is looked into only in cases where results or plots seem dubious to find the reason (bug in the code or conceptual error).

Your finished report has to be uploaded to the moodle **only by one person per group**. The submission function will be activated approximately one week before the deadline. **Don't forget to click 'Submit' in the end!**