Data Analysis and Model Classification
Mini-project 2: Description and Assignment

Fumiaki Iwane Ping-Keng Jao Bastien Orset Julien Rechenmann
Ricardo Chavarriaga José del R. Millán

November 19, 2018

Objective

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

In this mini-project 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 second 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.

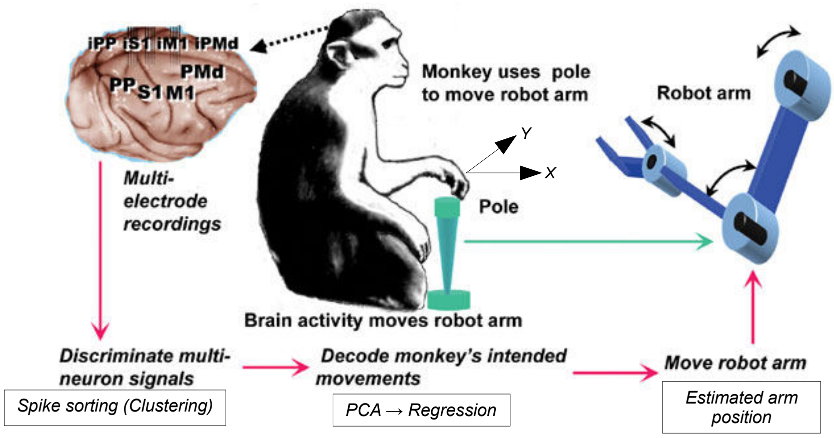

Figure 1: Experimental setup. Steps for data analysis are also shown.

The dataset *Data.mat* contains already the computed firing rates of all identified neurons. Here we denote $Data_t$ as t^{th} row (sample). Each row contains activities of 48 neurons in a 1-second window with 20Hz (50 ms) sampling rate. As a result, each row has 48 (neurons) $\times$ 20 (time sample) = 960 features. The next row is simply shifted by 50 ms. See the table 1 for a graphical representation.

Apart from the firing rates, the trajectories of arm movement are denoted as $PosX_t$ and $PosY_t$ and are also recorded during the experiment. The data for arm trajectories **PosX** and **PosY** is used as ground

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.

truth to design a regressor model where the input is neuronal activity. In other words, the firing rates $Data_t$ will be compared with the actual trajectories $PosX_t$ and $PosY_t$ to directly predict the trajectories from brain signals. 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: (Window) Samples × Features (48 neurons × 20 bins (50-ms)) 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 from 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 performance. Put the exact model description and the obtained performance (which metric?) in the report!