**Problem Set #10 (Classification & optimization)**

**28 points (3-4 hrs) 🡪 counts for 7.5 points on your final grade.**

**Due: Monday, 3/31/2025 (11.59 am)**

1. **Published literature (14 points; 2 hrs)***.* Listed below are papers that apply classification and clustering algorithms to problems in the brain sciences. Pick **any two** papers from this list (consider reading all their abstracts before choosing two), and answer the following questions for each. Answer briefly, i.e., use a sentence or three for each part.
2. (2 points) (10 min) What was the goal of the paper?
3. (2.5 points) (15 min) How did they choose to answer this question using clustering/classification methods? That is, how did they set up the experiments and design the analysis to address the goal?
4. (2.5 points) (15 min) What were some of the key aspects of the methods (selection of features, data organization, etc)?

***List of papers:***

* Behavior/cognition: Mandera 2015, Douglas 2011
* fMRI: Ball 2014;
* Single cell responses: Semmelhack 2014

**2 (6 points; 30 min – 1 hr).** The AIC value we obtained **in the lecture video/slides** (with commands reproduced below) is different from the AIC value produced by the **function fitSigmoid\_QMbb** (see below). Can you explain why? To answer this, proceed step-by-step as follows:

1. (1 point) Run the provided commands and code to compare AICs.
2. (3 points) Is it true that different formulae are used in the two cases? If so, can you fix the problem and produce the same AIC value? If not, what else explains the difference? (Hint: “help nlinfit”, check out details in both codes).
3. (2 points) Re-run both cases with the number of iterations being 2000. (Hint: “help nlinfit”). Does increasing the number of iterations make the AIC values in the two cases same (or more similar than before)?

Obtaining aic using commands used **in the lecture video/slides**

load ps10a\_data

[beta,res] = nlinfit(x,means1,@sigmoid,beta01);

nObs = length(x);

aic = 2\*4 + nObs\*(log((sum(res.^2))/nObs))

***vs.***

Obtaining aic using the **function**

[beta,stats1]= fitSigmoid\_QMbb(x,means1);

stats1.aic

**3 (5 points; 20 min).** From ps10a\_data.mat,

1. (2 points; 10 min) Use means2, and determine whether the sigmoidal model or the linear model is a better choice. Why?

*Hint. Use the same functions we used in class: fitSigmoid\_QMbb.m and fitLinear\_QMbb.m*

1. (3 points; 10 min) How confident are you of your answer, given your response to question 1? Discuss in a few sentences. (Hint: think about factors that affect AIC value)

**4. (3 points; 10 min)** Under what SPECIFIC condition is it true that the BIC penalizes model complexity more than the AIC? (Hint: see slides for formulas. Your focus should be on the values of N.)

**Notes**: Data, functions and papers for this problem set are all available as part of the zipped PS10 folder.