EEE 482/582: Computational Neuroscience

Homework 3

(Due 16/04/2019, 17:00)

Instructions:

1. Prepare a report (including your answers/plots) to be uploaded on Moodle.
2. The report should be typeset (for lengthy derivations, the solution can be scanned and embedded into the report).
3. Show all the steps of your work clearly.
4. Unclear presentation of results will be penalized heavily.
5. No partial credits to unjustified answers.
6. Use Matlab or Python for computations.
7. Return all Matlab/Python code that you wrote in a single file.
8. Code should be commented, code for different HW questions should be clearly separated.
9. The code file should NOT return an error during runtime.
10. If the code returns an error at any point, the remaining part of your code will not be evaluated (i.e., 0 points).

Question	Points	Your Score
Q1	20	
Q2	30	
Q3	30	
Q4	20	
TOTAL	100	

Question 1. [20 points]

Responses of a V1 neuron in an awake monkey are provided in the file `hw3_data1.mat`. This file contains two variables `resp1` and `resp2` that represent the response levels during a passive fixation condition and those during a memory task, respectively. Answer the questions below.

a) Using ordinary least-squares (OLS) method, fit a linear model $y = a \cdot x + b$, where the independent variable (i.e., regressor) is `resp1` and the dependent variable (i.e., output) is `resp2`. Show the original data on a scatter plot, and show the model fit with a line on the same graph. Report the estimated model parameters on the graph. Find the explained variance, unexplained variance, and the coefficient of determination (R^2) for the linear model. Compare R with Pearson's correlation coefficient between `resp1` and `resp2`.

b) Using OLS, fit a linearized second-order model $y = a \cdot x^2 + b \cdot x + c$, where `resp1` is the regressor and `resp2` is the output. Show the original data on a scatter plot, and show the model fit with a curve on the same graph. Report the estimated model parameters on the graph. Find the explained variance, unexplained variance, and the coefficient of determination (R^2). Compare R with Spearman's correlation coefficient between `resp1` and `resp2`.

c) Fit a parametric nonlinear model $y = a \cdot x^n + b$, where `resp1` is the regressor and `resp2` is the output. Use `lsqcurvefit` function to find the model parameters, starting the iterative algorithm at two different initial values: $\{a, n, b\} = \{1, 1, 0\}$ and $\{10, 7, 100\}$. Show the original data on a scatter plot, and show the two model fits with separate curves on the same graph. Report the estimated model parameters on the graph. Find the explained variance, unexplained variance, and the coefficient of determination (R^2) for both models. Which model performs better and why?

d) Fit a nonparametric nonlinear model using nearest-neighbor regression, where `resp1` is the regressor and `resp2` is the output. Show the original data on a scatter plot, and show the model fit on the same graph. Find the explained variance, unexplained variance, and the coefficient of determination (R^2). Interpret the meaning of your R^2 measurement.

Question 2. [30 points]

A two-alternative forced choice (2AFC) experiment is conducted, in which a subject views two stimulus arrays of potentially different intensities side by side. In each trial, the subjects is assigned the task to determine which array contains the target stimulus (only one of the arrays contain the target in each trial). The probability of giving a correct answer on each trial (i.e., the psychometric function) is given by:

$$p_c(I) = \frac{1}{2} + \frac{1}{2}\Phi(I; \mu, \sigma) \tag{1}$$

where $\Phi(I; \mu, \sigma)$ is the CDF of a Gaussian random variable with mean μ and standard deviation σ, evaluated at stimulus intensity I. Answer the questions below.

a) Plot the pscyhometric functions for $\{\mu, \sigma\}$ equal to $\{6, 3\}$ and equal to $\{3, 4\}$, for $I \in [1\ 10]$. Describe and interpret the differences between the two functions. Is the range of $p_c(I)$ appropriate based on the description of the 2AFC experiment?

b) Write a function `[C,E] = simpsych(mu,sigma,I,T)` that takes two vectors `I` and `T` of the same length, containing the stimulus intensities and number of trials for each intensity, respectively. The function should simulate random draws from $p_c(I)$. Outputs are a vector `C` that contains the number of trials correct out of `T` at each stimulus intensity `I`, and a matrix `E` that contains the trial result at each stimulus intensity and each of `T` trials at that intensity. Show a scatter plot of `C/T` versus `I`, and a curve plot of $p_c(I)$ on the same graph, assuming `T=ones(1,7)*100`, $I \in [1 : 1 : 7]$, $\mu = 5$ (`mu`), and $\sigma = 1.5$ (`sigma`).

c) Write a function `nll = nloglik(pp,I,T,C)` that returns the negative of the log-likelihood of parameters `pp.mu, pp.sigma` given an experimental dataset of `I,T,C`. Show a contour plot (using 50 contours) of the negative log-likelihood of the dataset generated in **part b** (i.e., using the `I,T,C` vectors from the previous part), for all pairs $\{\mu_{est}, \sigma_{est}\}$ where $\mu_{est} \in [2 : 0.1 : 8]$ and $\sigma_{est} \in [0.5 : 0.1 : 4.5]$. Determine the best fitting parameters $\{\mu_{est}, \sigma_{est}\}$ by visual inspection.

d) Using the function `fminsearch`, find a more precise estimate of $\{\mu_{est}, \sigma_{est}\}$ that minimizes the function `nloglik`. **Hints:** You can call `fminsearch` with an initial starting point of `{2,2}` for $\{\mu_{est}, \sigma_{est}\}$. You are strongly recommended to use the `fit.m` package (the `fminsearch` interface coded by Geoff Boynton) that is available on Moodle. Be careful when calling this function, since there is also a native `fit.m` function in Matlab.

e) Determine confidence intervals for the parameter estimates using the bootstrap technique. For each stimulus intensity, generate bootstrap samples by resampling the 100 trials of that intensity in the original data (i.e., resample `E`). A set of bootstrap samples for all stimulus intensities consititutes a resampled dataset; refit the model to this dataset using `fminsearch`. Perform 200 bootstrap iterations, and plot the histograms of $\{\mu_{est}, \sigma_{est}\}$. Find the 95% confidence intervals.

Question 3. [30 points]

Blood-oxygen level dependent (BOLD) responses of a neural population in human visual cortex are provided in the file `hw3_data2.mat`. This file contains a variable `Yn` that represents 1000 response samples. There is another variable `Xn` that represent 100 regressors that may explain the responses. For all parts, the proportion of explained variance (R^2) should be calculated as the square of Pearson's correlation coefficient between measured and predicted responses. Answer the questions below.

a) Use the ridge regression method to fit regularized linear models to predict noisy BOLD responses as a weighted sum of given regressors. Perform 10-fold cross-validation to tune the ridge parameter ($\lambda \in [0\ 10^{12}]$) based on model performance. (Hint: Vary the ridge parameters logarithmically.) Note that for $\lambda = 0$, the model obtained with ridge regression is equivalent to the OLS solution. For each cross-validation fold, do a three-way split of the data: select a validation set of 100 contiguous samples, a testing set of 100 samples (that immediately precede the validation set assuming circular symmetry), and a training set of length 800 samples. Fit a separate model for each λ using the training set. Find R^2 of each model on the testing set. Separately estimate R^2 of each model on the validation set. Plot the average R^2 across cross-validation folds, measured on the testing set as a function of λ. Find the optimal ridge parameter λ_{opt} that maximizes average R^2. Find the model performance by calculating the average R^2 across cross-validation folds, measured on the validation set for λ_{opt}. Plot R^2 curves obtained on testing and validation data for all λ values. Interpret your results.

b) Determine confidence intervals for parameters of the OLS model from **part a** (i.e., the model obtained for $\lambda = 0$). Generate bootstrap samples from the 1000 samples in the original data (resample both the regressors and the responses the same way). Perform 500 bootstrap iterations, and refit a separate model at each iteration. Plot the mean and 95% confidence intervals of the parameters in the same graph. Identify and label on your plots, the model regressors which have weights that are significantly different than 0 (at a significance level of $p < 0.05$).

c) Determine confidence intervals for parameters of the regularized linear model from **part a** (i.e., the model obtained for λ_{opt}). Generate bootstrap samples from the 1000 samples in the original data (resample both the regressors and the responses the same way). Perform 500 bootstrap iterations, and refit a separate model at each iteration using λ_{opt} found in **part a**. Plot the mean and 95% confidence intervals of the parameters in the same graph. Identify and label on your plots, the model regressors which have weights that are significantly different than 0 (at a significance level of $p < 0.05$). Compare the results to those in **part b**.

Question 4. [20 points]

A series of neural response measurements are provided in the file `hw3_data3.mat`. Answer the questions below to examine the relationship between these measurements. Provide plots whenever possible.

a) Responses from two separate populations of neurons are stored in the variables `pop1` and `pop2`. We would like to examine whether the mean responses of the two populations are significantly different. The first population contains 7 neurons, whereas the second population contains 5 neurons. Using the bootstrap technique (10000 iterations), find the two-tailed p-value for the null hypothesis that the two datasets follow the same distribution. (Hint: If the two datasets come from a common distribution, is there any need to separate them?)

b) BOLD responses recorded in two voxels in the human brain are stored in the variables `vox1` and `vox2`. We would like to examine whether the voxel responses are similar to each other, by calculating their correlation. Using the bootstrap technique (10000 iterations), find the mean and 95% confidence interval of the correlation. Find the percentile of the bootstrap distribution, corresponding to a correlation value of 0. (Hint: Should you resample `vox1` and `vox2` independently or identically?)

c) Note that estimation of confidence intervals and hypothesis testing are dual problems. For the dataset examined in **part b**, use bootstrapping (10000 iterations) to simulate the distribution of the null hypothesis that two voxel responses have zero correlation. Find the one-tailed p-value for the two voxel responses having zero or negative correlation. Compare this to the result in **part b**. (Hint: Resample the datasets to break apart the correlation between them.)

d) The average BOLD responses in a face-selective region of the human brain have been recorded in two separate experiments. The responses of this region to building images (1st experiment) and face images (2nd experiment) are stored in the variables `building` and `face` for 20 subjects. Assume that the same subject population was recruited in both experiments. Use bootstrapping (10000 iterations) to calculate the two-tailed p-value for the null hypothesis that there is no difference between the building and face responses.

e) Repeat the exercise in **part d**, but this time assuming that the subject populations recruited for the two experiments are distinct. Use bootstrapping (10000 iterations) to calculate the two-tailed p-value for the null hypothesis that there is no difference between the building and face responses.