EEE 482/582: Computational Neuroscience

Homework 1

(Due 19/02/2019, 17:00)

Instructions:

1. Prepare a report (including your answers/plots) to be uploaded on Moodle.
2. The report should be typeset (no handwriting allowed except for lengthy derivations, which may 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	25	
Q2	25	
Q3	25	
Q4	25	
TOTAL	100	

Question 1. **[25 points]** Assume that a neural population computes weighted linear combinations of its input x, characterized by a system of equations $Ax = b$. Here A is the transfer function and b is the output vector.

A single output measurement is recorded, given by $\begin{pmatrix} 1 & 0 & -1 & 2 \\ 2 & 1 & -1 & 5 \\ 3 & 3 & 0 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 9 \end{pmatrix}$.

Answer the questions below. Derive the results by hand first and then confirm them on the computer.

a) Find all solutions x_n to $Ax = 0$.

b) Find a particular solution x_p to $Ax = b$.

c) Find all solutions to $Ax = b$.

e) Find the pseudo-inverse of A.

e) Find the sparsest solution to $Ax = b$ (the solution with the least number of non-zero entries).

f) Find the least-norm solution to $Ax = b$ (the solution with the minimum Euclidean norm).

Question 2. [25 points]

We record the aggragated responses of a population of 100,000 neurons (rough number of neurons contained in a volumetric-pixel of fMRI acquisitions). 30 separate recordings of the aggragate response X are given in `Xdata.mat`. Answer the questions below. Derive the results on the computer.

a) Find the sample mean, and the sample median.

b) Find the sample standard deviation, and the sample inter-quartile range.

c) Compute the histogram of the data for the value range $70 \leq X \leq 130$. Plot the histograms for n = 3, 6, 12 bins.

d) Determine whether X is normally distributed by running the following code on samples of X. (This is called the Normal Quantile Plot)

```
y = sort(x);
n = length(x);
f = ((1:n)-3/8)/(n+1/4);
q = 4.91*(f.^0.14 - (1-f).^0.14);
figure(1);clf;plot(q,y,'*-');grid;
```

e) Using bootstrap resampling with N = 1000 iterations, determine the sample mean and its standard error. Find the 95% confidence interval of the mean. Plot a histogram of the bootstrap distribution for n = 50 bins.

f) Using bootstrap resampling with N = 1000 iterations, determine the sample standard deviation and its standard error. Find the 95% confidence interval of the standard deviation. Plot a histogram of the bootstrap distribution for n = 50 bins.

g) Repeat **parts e and f** using jackknife resampling. Note that for jackknife, you resample by leaving one-sample-out at each iteration, so N = 30. Plot histograms for n = 5 bins.

Question 3. [25 points]

'Reverse inference' is a common, albeit poorly exercised method in neuroscience. It refers to the practice of inferring that a cognitive process is engaged on the basis of activation in some brain area. For example, if Broca's area was found to be activated in some task, researchers might infer that the subjects were using language. After a comprehensive search of the literature, we find that Broca's area was reported to be activated in 103 out of 869 fMRI tasks involving engagement of language, but this area was also active in 199 out of 2353 tasks not involving language.

a) Assume that the conditional probability of activation given language and activation given no language, each follow a Bernoulli distribution (i.e., active with some probability p, or not with probability $1-p$). Compute the likelihoods of observed frequencies of activation in literature, as functions of the possible values of their respective Bernoulli probability parameters $p = x_l$ and $p = x_{nl}$. Compute these functions at the values x =[0:.001:1] and plot them as separate bar charts.

b) Find the values of x_l and x_{nl} that maximize their respective discretized likelihood functions.

c) Using the likelihood functions computed for discrete x, compute and plot the discrete posterior distributions $P(X|data)$ and the associated cumulative distributions $P(X \leq x|data)$ for both processes (language and no language cases). To do this, assume a uniform prior $P(x) \propto 1$ and note that it will be necessary to compute (rather than ignore) the normalizing constant for Bayes' rule. Use the cumulative distributions to compute (discrete approximations to) upper and lower 95% confidence bounds on each proportion ($x_{l,nl}$).

d) Consider the joint posterior distribution $P(X_l, X_{nl}|data)$ over x_l and x_{nl}, the Bernoulli probability parameters for the language and non-language contrasts. Given that these two frequencies are independent, the (discrete) joint distribution is given by the outer product of the two marginals. Plot it (with `imagesc`). Compute (by summing the appropriate entries in the joint distribution) the posterior probabilities that $P(X_l > X_{nl}|data)$ and conversely that $P(X_l \leq X_{nl}|data)$.

e) Using the estimates from **part b** as the relevant conditional probabilities, and assuming the prior that a contrast engages language, $P(language) = 0.5$, compute the probability $P(language|activation)$ that observing activation in this area implies engagement of language processes. Is the critique on 'reverse inference' correct? How confident should you be in implicating language if you observe activity in Broca's area?

Question 4. [25 points]

Write a function `samples = ndRandn(mean, cov, num)` that generates a set of samples drawn from a multidimensional Gaussian distribution with the specified mean (an N-vector) and covariance (an NxN matrix). The parameter `num` should be optional (defaulting to 1) and should specify the number of samples to return. The returned value should be a matrix with num rows each containing a sample of N elements.

a) Your function should use the MATLAB function `randn` to generate samples from an N-dimensional Gaussian with identity covariance matrix, and then modify these appropriately. (Hint: Recall that the sample covariance is $1/N \cdot X^T X$. If $1/N \cdot X^T X$ is I then $(XY)^T(XY)$ has sample covariance $1/N \cdot Y^T Y$. Therefore you need to multiply samples X by a matrix Y such that $Y^T Y =$ `cov`. You can find such a matrix using the SVD of `cov`).

b) Test your function by plotting 1000 samples of a 2-dimensional Gaussian (choose an arbitrary nonzero mean and nonzero covariance). Measure the sample mean and covariance of your data points, comparing to the values that you requested when calling the function. Plot an ellipse on top of the scatterplot that traces out points that are two standard deviations away from the mean, according to the covariance matrix. Does this ellipse capture the shape of the data?