ECE/BME 6790 - Information Processing Models in Neural Systems

Spring 2019 - Homework 3

Due 02.12.2019

Remember that presentation and readability matter!

1. There are very few quantities in natural signals or systems that are truly independent, so we need to know how to recognize statistical dependence when we see it.

   Download the file `HW_corr_data.mat` from the course website, and use the `load` command to load it into Matlab. The variable `data1` contains 10000 realizations of a pair of Bernoulli random variables x_1, y_1. The variable `data2` contains 10000 realizations of another pair of Bernoulli random variables x_2, y_2.

   (a) Estimate the marginal distributions for x_1, y_1, x_2 and y_2. Display the results using a bar or histogram plot.

   (b) One of these pairs is independent and one has dependency between the variables. Examine the data and determine which pair of variables is most likely independent. Present an argument (using figures!) to support your decision.

2. Following up on the previous problem, this problem will get you to generate data with significant correlations and then examine it from the viewpoints of geometry and linear algebra.

   This problem will use Matlab to generate and examine pairs of random variables from a Gaussian distribution using the built-in function `randn`.

   (a) Generate a dataset with dependent pairs of Gaussian random variables. Generate this dataset by first generating 1000 trials of independent Gaussian pairs, where the first RV has variance 1 and the second has variance 16 (i.e., the covariance is $K = [1,\ 0; 0,\ 16]$). Now multiply each pair by the matrix:

   $$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

   For this problem, use $\theta = -\pi/4$.

   Make a scatter plot of the data using the `plot` command with the '.' linestyle option (i.e., each trial should be a single dot in a 2-D plane). Make sure that the axis are the same in both plots so you can compare the two datasets. You may also want to use the `axis square` option to make the perspective of the plot correct. Also plot a histogram of both marginal distributions with the Gaussian PDF superimposed (see the `normpdf` function) to verify their Gaussianity. You may have to scale the PDF to compare it's shape to the histogram if the histogram has not been normalized (do not worry that this scaling will make it an ill-formed PDF).

   (b) Analytically determine the covariance matrix forthis dataset and confirm this calculation by using the built-in function `cov` to estimate the covariance matrix.

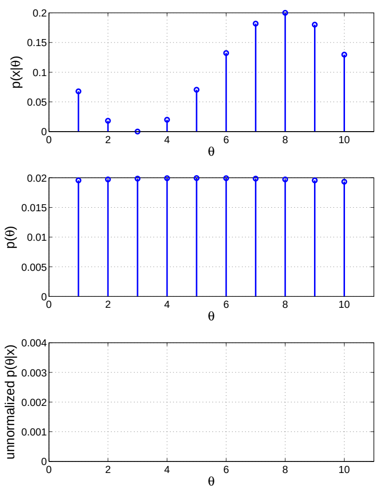
   (c) Use the built-in function `eig` to calculate the eigenvectors and eigenvalues of the estimated covariance matrix. Superimposed on the scatter plot of the data, plot vectors with a direction given by the eigenvectors and length given by the square root of the eigenvalues. Comment on how we should interpret the eigenvectors of the covariance matrix. Recall from our earlier discussions that this is the same as calculating the PCA vectors for the dataset.

   (d) Create another dataset as above but with $\theta = \pi/4$, and merge your two datasets together (i.e., concatenate the datasets into one big larger dataset, which is **very different** from adding them). Make the same scatter plot and marginal histograms with the Gaussian PDF superimposed. Do you think this data still Gaussian distributed?

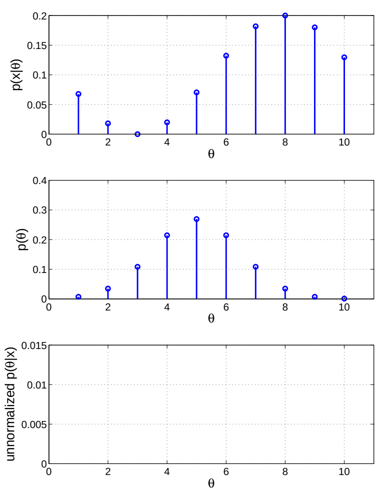
(e) Estimate the covariance matrix of this data and calculate it's eigenvectors and eigenvalues. Plot these PCA vectors (as above) superimposed on the scatter plot of the data. How does your interpretation of the eigenstructure of the covariance matrix change in this case?

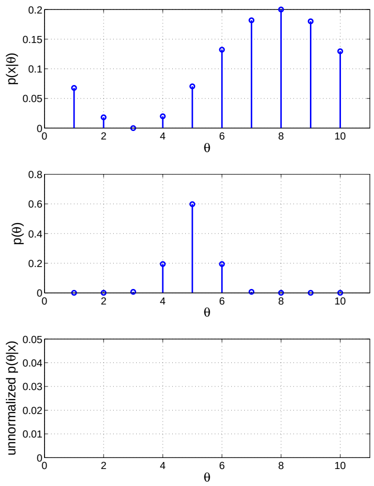
3. Consider a sensory system trying to determine the value of a scalar stimulus parameter θ that can take integer values $\theta = 1, \ldots, 10$. A neural system observes the stimulus $\boldsymbol{x}$ and the response of each cell encodes the likelihood function for that stimulus given a different parameter, $p(\boldsymbol{x}|\theta = i)$ for $i = 1, \ldots, 10$. For sections (a)-(c) of this question, sketch the (unnormalized) posterior distribution for θ under the given prior $p(\theta)$. In other words, don't worry about calculating the partition function to make sure the posterior distribution sums to one. You can use the figures provided to draw in if you find it easier, but you don't have to. For each section, also report the maximum likelihood estimate (MLE) and maximim a posteriori (MAP) estimate for θ.

(a)



(b)



(c) 

(d) In words, explain how and why the particular changes in $p(\theta)$ in sections (a)-(c) effect the MLE and MAP estimates of θ.