EE278 Statistical Signal Processing **Stanford, Autumn 2023**

Homework 6

Due: Thursday, November 16, 2023, 1:00 pm on Gradescope

Please upload your answers timely to Gradescope. Start a new page for every problem. For the programming/simulation questions you can use any reasonable programming language. Comment your source code and include the code and a brief overall explanation with your answers.

1. (**8 pts**) Exercise 3.9 in text.

   Let X and Y be jointly Gaussian with means m_X, m_Y, variances σ_X^2, σ_Y^2, and normalized covariance ρ. Find the conditional density $f_{X|Y}(x|y)$.

2. (**15 pts**) Suppose $\mathbf{X}$ and $\mathbf{Y}$ are two n-dimensional random vectors. They are said to be *uncorrelated* if the ith component of $\mathbf{X}$ and the jth component of $\mathbf{Y}$ are uncorrelated for all i, j. They are said to be independent if for every $\mathbf{x}, \mathbf{y}$,

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y}).$$

   a) (**3 pts**) Does the above definition of uncorrelatedness above say anything about whether the components of $\mathbf{X}$ are uncorrelated of each other? Does the above definition of independence say anything about whether the components of $\mathbf{X}$ are independent of each other? Explain.

   b) (**4 pts**) Assume $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated. Compute the covariance matrix of $\mathbf{Z} := \mathbf{X} + \mathbf{Y}$ in terms of the covariance matrices of $\mathbf{X}$ and of $\mathbf{Y}$.

   c) (**4 pts**) In general, compute the covariance matrix of $\mathbf{Z}$ in terms of the covariance matrices of $\mathbf{X}$ and of $\mathbf{Y}$ and something else. Define this "something else" appropriately in terms of a matrix.

   d) (**4 pts**) Suppose $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian, i.e. the random vector

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$$

   is jointly Gaussian. If $\mathbf{X}$ and $\mathbf{Y}$ are independent, does it imply that they are uncorrelated? If $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated, does it imply that they are independent. Fully justify your answers. You can assume that K_X and K_Y are invertible for this part of the question.

3. Multiple Choice Exam **(8 pts.)**

You are taking a multiple choice exam. Question number 1 allows for two possible answers A and B. According to your first impression, answer A is correct with probability $1/4$ and answer B is correct with probability $3/4$. You would like to maximize your chance of giving the correct answer and you decide to have a look at what your neighbors on the left and right have to say. The neighbor on the left has answered A. He is an excellent student who has a record of being correct 90% of the time when asked a binary question. The neighbor on the right has answered B. He is a weaker student who is correct 70% of the time.

(a) **(4 pts.)** You decide to use your first impression as a prior and to consider your two neighbors' answers as your observations. Formulate the problem as a decoding problem.

(b) **(4 pts.)** What is your answer?

4. Exercise 8.9 in text **(10 pts.)**

   A disease has two strains, 0 and 1, which occur with a priori probabilities p_0 and $p_1 = 1 - p_0$ respectively.

   a) **(2 pts.)** Initially, a rather noisy test was developed to find which strain is present for patients with the disease. The output of the test is the sample value y_1 of a rv Y_1. Given strain $0(X=0), Y_1 = 5 + Z_1$, and given strain $1(X=1), Y_1 = 1 + Z_1$. The measurement noise Z_1 is independent of X and is Gaussian, $Z_1 \sim \mathcal{N}\left(0, \sigma^2\right)$. Give the MAP decision rule, i.e., determine the set of observations y_1 for which the decision is $\hat{x} = 1$. Give $\Pr\{e \mid X = 0\}$ and $\Pr\{e \mid X = 1\}$ in terms of the function $Q(x)$.

   b) **(2 pts.)** A budding medical researcher determines that the test is making too many errors. A new measurement procedure is devised with two observation rv Y_1 and $Y_2 . Y_1$ is the same as in (a). Y_2, under hypothesis 0 , is given by $Y_2 = 5 + Z_1 + Z_2$, and, under hypothesis 1 , is given by $Y_2 = 1 + Z_1 + Z_2$. Assume that Z_2 is independent of both Z_1 and X, and that $Z_2 \sim \mathcal{N}\left(0, \sigma^2\right)$. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2), and find $\Pr\{e \mid X = 0\}$ and $\Pr\{e \mid X = 1\}$. Hint: Find $f_{Y_2 \mid Y_1, X}\left(y_2 \mid y_1, 0\right)$ and $\mathrm{f}_{Y_2 \mid Y_1, X}\left(y_2 \mid y_1, 1\right)$.

   c) **(2 pts.)** Explain in laymen's terms why the medical researcher should learn more about probability.

   d) **(2 pts.)** Now suppose that Z_2, in (b), is uniformly distributed between 0 and 1 rather than being Gaussian. We are still given that Z_2 is independent of both Z_1 and X. Find the MAP decision rule for $\hat{x}$ in terms of the joint observation (y_1, y_2) and find $\Pr(e \mid X = 0)$ and $\Pr(e \mid X = 1)$.

   e) **(2 pts.)** Finally, suppose that Z_1 is also uniformly distributed between 0 and 1 . Again find the MAP decision rule and error probabilities.

5. **(9 pts.)** This problem will explore optimal classification based on the features selected via principal component analysis (PCA). Download the MNIST handwritten digit dataset

from the course website http://web.stanford.edu/class/ee278/homeworks/hw6-data.zip. The folders `train0` and `train2` contain the same set of images we used in Homework 5. We will use these images to train a classifier that can distinguish between the digits "0" and "2". We also have added two test sets (each with 500 images) of the handwritten digit "0" and of the digit "2" in the folders `test0` and `test2`.

a) **(3 pts.)** As in Homework 2, consider each image as vector $\boldsymbol{X}_i \in \mathbb{R}^{784}$. Combine all the training images (in folders `train0` and `train2`) and generate an estimator of the covariance matrix. Compute the first 20 eigenvectors $\boldsymbol{U}_i$ ($1 \leq i \leq 20$) corresponding to the largest eigenvalues. Now project each training image $\boldsymbol{X}_i$ onto the new set of basis vectors $\boldsymbol{U}_i$. The result, denoted $\tilde{\boldsymbol{X}}_i \in \mathbb{R}^{20}$, is a lower dimensional feature vector that we will use to represent the data.

b) **(3 pts.)** Estimate the mean and covariance matrix of the training vectors $\tilde{\boldsymbol{X}}_i$ corresponding to the digit "0" and the mean and covariance matrix corresponding to the digit "2". Suppose that each digit-0 (resp. digit-2) image $\tilde{\boldsymbol{X}}_i$ is independently drawn from a jointly Gaussian distribution $\mathbb{P}_0$ (resp. ,$\mathbb{P}_2$). Propose a maximum likelihood detector that classifies a given image as a "0" or "2", i.e. assuming an equal prior. (This is called Gaussian discriminative analysis in machine learning.)

c) **(3 pts.)** Run your classifier on the test dataset in the folders `test0` and `test2`. Report the empirical error rates.