Homework 1

Due on Wednesday, $\boxed{\textbf{4/1/2019}}$, at $\boxed{\textbf{12:00 noon}}$ in class.
Be noted that late homework will **NOT** be accepted!!

Grading Policy

- Please work all the five problems. Each subproblem takes up to 10 points. The total of the to-be-graded three problems is 110 points.
- You need to provide detailed proof or derivations that lead to your final answers. Add comments to your programming codes. You will receive no points if your answers are not supported by detailed explanations. So, do not skip steps.
- Partial points will be credited to you when a wrong answer is accompanied by detailed correct reasoning.
- Discussions are encouraged. But plagiarism is strictly prohibited. **You will fail the course as penalty for plagiarism!!**

Part I -- Reading assignment

Please read Section 1.1~Section 1.5, Chapter 2 (except Sec. 2.3.8), and Chapter 3 of the textbook. You don't need to turn in anything for this part. I didn't touch Section 1.5 during class lectures, but the materials therein will be required in later chapters. For those of you who haven't taken Stochastic Processes or Detection & Estimation, please familiarize yourself to Section 1.5 by self-study.

Part II -- Problem assignment

1. We flip a coin M times, with each toss statistically independent of all others and each with unknown probability p of heads. Let N be the number of heads that is observed.
   (a) Find the maximum likelihood (ML) estimator $\hat{p}_{ML}$ of p from the knowledge of $N = n$.
   (b) Evaluate the bias $E[\hat{p}_{ML}] - p$ and the mean square error $E[(\hat{p}_{ML} - p)^2]$ of the ML estimator.
   (c) Please show that $\lim_{M\to\infty} P\left[|\hat{p}_{ML} - p| > \epsilon\right] = 0$ for any positive ϵ. (Hint: Use Chebyshev inequality.)

2. We are given two coins, Coin 1 with bias $1/2$ and Coin 2 with random bias P that is uniformly distributed in $[0, 1]$. We pick one coin at random and flip it twice independently to generate the random variables X_i, for $i = 1, 2$. Specifically, $X_i = 1$ if the ith coin flip is heads and 0 if it is tails, for $i = 1, 2$. Note that X_1 and X_2 are only conditionally independent given the value of the coin bias.

   (a) Specify the maximum *a* posteriori probability (MAP) decision rule $D(X_1, X_2) \in \{1, 2\}$ for deciding whether the coin you selected is Coin 1 or Coin 2.
   (b) Find the corresponding probability of decision error.

3. Let X_1, X_2 and X_3 be jointly Gaussian random variables with the following properties: $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ are independent, $X_3 \sim \mathcal{N}(0, 10)$, $E[X_1X_3] = -1$, and $E[X_2X_3] = 2$.

(a) Show that the covariance matrix of $[X_1,\ X_2]^T$ conditioning on X_3 takes the form

$$\mathbf{K} = \mathbf{I} + \alpha \cdot \mathbf{u}\mathbf{u}^T, \tag{1}$$

where $\mathbf{I}$ is the 2×2 identity matrix, $\mathbf{u} \in \mathbb{R}^{2\times 1}$ with $||\mathbf{u}|| = 1$, and $\alpha \in \mathbb{R}$ is a constant. What are the values of the vector $\mathbf{u}$ and the constant α in this problem?

(b) Show that in general $\mathbf{u}$ is an eigenvector of any covariance matrix that takes the form in equation (1).

4. Suppose you are hired by NCTU to predict whether it will rain for the upcoming yearly Mei-Chu Game. Checking past data, you are aware of that the chance of rain is of 20%. You model this by a random variable R with

$$P(R = 1) = 0.2, \quad P(R = 0) = 0.8,$$

where $R = 1$ means that it rains on the game day and $R = 0$ otherwise. In addition to the history (i.e. $P(R = 1) = 0.2, \quad P(R = 0) = 0.8,$) of the weather on the day of the game, you decide to adopt the forecast from the Central Weather Bureau (CWB) to help you predict the weather on the game day. Analyzing the data from past forecasts of CWB, you find that CWB has correct forecast 70% of the time, which can be modeled as

$$P(W = 1|R = 1) = P(W = 0|R = 0) = 0.7,$$

where $W \in \{1, 0\}$ denotes the forecast result of CWB.

(a) Please devise a prediction strategy, after having learned the forecast result W from CWB for the upcoming Mei-Chu game, for the weather (rain or not rain) on the Mei-Chu game day that minimizes the prediction error probability (i.e., devise the MAP decision rule). Find the corresponding probability of error prediction.

In order to enhance the prediction accuracy, you decide to incorporate the relative humidity, which is modeled by a random variable H, of the air provided by CWB for the prediction. Assuming that H and W are conditionally independent given R, which means

$$f_{H|W,R}(h|w, r) = f_{H|R}(h|r),$$

where $f_{H|W,R}$ is the conditional pdf of H given W, R, and $f_{H|R}$ is the conditional pdf given R. The data from CWB establishes that conditioned on $R = 1$, H is uniformly distributed between 0.5 and 0.7, whereas conditioned on $R = 0$, H is uniformly distributed between 0.1 and 0.6.

(b) Please devise your forecast strategy if $H = 0.55$ and the CWB predicts rain. Justify your answer by comparing the posterior probabilities.

5. It is known that a certain type of heart disease is highly correlated with the patient's gender, type of chest pain, and cholesterol level. In this problem, you will implement the maximum a posteriori probability (MAP) detector of the heart disease for 50 patients, with their gender, type of chest pain, and cholesterol level provided in the test data set `heartdatasetTesting.mat` (which can be found as a separate file together with this homework assignment). The mathematical formulation is as follows:

Let H, S, C be discrete random variables such that

$$H = \begin{cases} 0, & \text{if the patient does not have a heart disease,} \\ 1, & \text{if the patient has a heart disease,} \end{cases}$$

$$S = \begin{cases} 1, & \text{if the patient is a female,} \\ 2, & \text{if the patient is a male,} \end{cases}$$

$$C = \begin{cases} 1, & \text{if the pain is typical angina,} \\ 2, & \text{if the pain is atypical angina,} \\ 3, & \text{if the pain is non-anginal pain,} \\ 4, & \text{if without chest pain symptoms.} \end{cases}$$

Let X be the continuous random variable denoting the patient's serum cholesterol in mg/dl.

In order to minimize the probability of detection error, you need to implement the MAP detection rule to determine whether a patient has the disease based on his/her gender, chest pain type, and cholesterol level (See Section 1.5 for details of MAP). More specifically, the MAP rule is

$$P[H = 1|S = s, C = c, X = x] \underset{\hat{H}=0}{\overset{\hat{H}=1}{\gtrless}} P[H = 0|S = s, C = c, X = x],$$

where $\hat{H}$ is the decision result, $s \in \{1, 2\}$ and $c \in \{1, 2, 3, 4\}$. The above MAP decision rule compares the posterior probabilities $P[H = h|S = s, C = c, X = x]$, $h \in \{0, 1\}$: if $P[H = 1|S = s, C = c, X = x] > P[H = 0|S = s, C = c, X = x]$, then the detector produces the decision of $\hat{H} = 1$; otherwise the decision goes to $\hat{H} = 0$.

To evaluate the posterior probabilities, you need to learn relevant likelihood functions and prior distribution from the training data set `heartdatasetTraining.mat`. This part has been written in the MATLAB code `heartDisease.m`. You need to extend the code to solve the following two problems.

(a) Assume that S, C, and X are conditionally independent given H. We further assume that $f_{X|H}(x|1) \sim \mathcal{N}(254, 2047)$, $f_{X|H}(x|0) \sim \mathcal{N}(245, 2182)$. From the given MATLAB code `heartDisease.m`, the values of $p_{S|H}[s|h]$ and $p_{C|H}[c|h]$ for all s, c, and $h \in \{0, 1\}$ can be empirically evaluated using the provided data, which are required when finding the posterior probability $P[H = h|S = s, C = c, X = x]$.
Please use the information provided in the code `heartDisease.m` and extend the code to produce the MAP detection result $\hat{H}(s, c, x)$ for the 50 patients with their $S = s$, $C = c$, $X = x$ given in the test data `heartdatasetTesting.mat`. In other words, you need to write a MATLAB code based upon `heartDisease.m` to implement the MAP detection, and output the detection result of the 50 patients. You have to clearly describe how you obtain the posterior probability in your written homework sheet and add corresponding comments in your code.

(b) Following part (a), please evaluate the detection error rate of the MAP detector by comparing your result in part (a) to the true H of each patient in the test data `heartdatasetTesting.mat`.