

PROBLEM SET #2

Due 5:00 P.M., October 6, 2020

Problem 1

Do problem 2.5 in your text.

Problem 2

Do problem 2.10 in your text. In addition do the following:

h) Are X and Y independent? Show this by determining the marginal distributions $f_X(x)$ and $f_Y(y)$.

Problem 3

	Number of Residents No Felony Conviction	Number of Residents with One or More Felony Convictions
Female Residents	359,881	172,742
Male Residents	340,454	40,854

You are visiting an island that is a prison colony for political prisoners from the dystopian planet Tralfamador. When a political prisoner receives a guilty verdict in court, (s)he is declared a felon and sent to the island along with their immediate family members. The felony conviction statistics for the current residents of the island are given in the table above. Consider the following outcomes for a resident of the island picked at random: (1) M = male resident; (2) F = female resident; (3) C = convicted felon; (4) I = no felony record. Using the data given and assuming that relative frequencies are equal to probabilities, calculate the following: $P(M)$, $P(F)$, $P(C)$, $P(I)$, $P(M \cap I)$, $P(F \cap I)$, $P(M \cap C)$, $P(F \cap C)$. Now using Bayes' rule answer the following questions. Be sure to check your answers using relative frequencies computed from the given data.

- What is the probability that a male resident of the island picked at random will be a convicted felon?
- What is the probability that a female resident of the island picked at random will be a convicted felon?
- What is the probability that an island resident picked at random is female, given that they are a convicted felon?

- d) What is the probability that an island resident picked at random is male, given that they are a convicted felon?

Problem 4

Do problem 3.9 in your text. This is the problem we kicked-off the course with and now you have the tools to find its solution. **HINT:** Treat it as a recursive least squares problem.

Problem 5

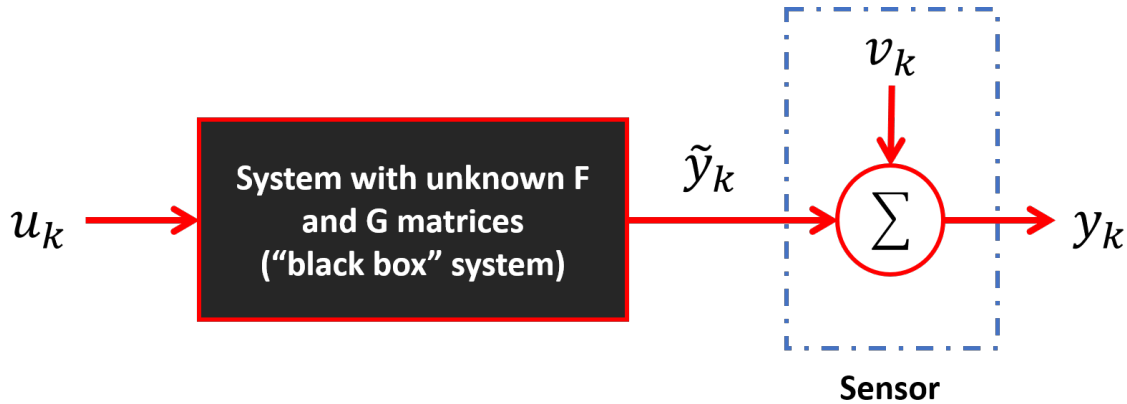


Figure 1: A system with an unknown mathematical model.

This problem demonstrates the utility and power of a simple, least squares estimator. You will use least squares to determine the mathematical model of an unknown dynamical system. The system can be an unknown electrical circuit, for example, which you can stimulate and then record its output (e.g., voltage-in-voltage out). Or alternately, it can be a vehicle whose equations of motion you do not know. So you apply control inputs (e.g., turn the vehicle right and left rapidly; accelerate and decelerate it, etc) and observe its kinematic state (e.g., position, velocity, acceleration, etc). Such problems are called *system identification* (or simply *sysid*) and they are a small part of what we know as *machine learning* today.

Download the data file `inoutdata.txt` found on the assignment page. This data contains the output from the unknown (or “blackbox”) system shown in Figure 1. The first column of the data is the time t_k , the second column is the input u_k and the third column is the output \tilde{y}_k (that is, y_k when $v_k = 0$). The system dynamics and observation models are:

$$\begin{aligned}\alpha_k &= F\alpha_{k-1} + Gu_{k-1} \\ y_k &= \tilde{y} + v_k = C\alpha_k + v_k\end{aligned}$$

Suppose you know that the matrices F , G and H have the following form:

$$F = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad C = [1 \quad 0]$$

Your job is to use least squares to estimate (or identify) the parameters or entries of F and G . That is, generate an optimal estimate of:

$$x = [a_1 \quad a_2 \quad b_1 \quad b_2]^T$$

You will proceed to do this as follows:

- a) Write out the measurement equation for your estimator in the form

$$\tilde{y} = Hx$$

where \tilde{y} is the error-free sensor measurement (i.e., case when $v = 0$). Clearly identifying the entries in \tilde{y} and H . This is by far the most challenging part of the problem. **HINT:** Try to relate y_k to $y_{k-1}, y_{k-2}, \dots, u_{k-1}, u_{k-2}, \dots$ and so using an autoregressive moving average (ARMA) model. If you do not know what an ARMA model is, look it up.¹

- b) Estimate the state vector using normal least squares.
- c) Note that error-free \tilde{y}_k is not available to you in practice. Instead you will use some sensor to measure an error corrupted version denoted y_k . Use the MATLAB command `randn` to simulate the sensing process as follows:

$$y_k = \tilde{y}_k + 0.1 * \text{randn}(N, 1)$$

where N is the number of data points in \tilde{y}_k you are using for the estimation process.

- d) Use the history of \tilde{y}_k you generated and u_k to generate an estimate of x .
- e) Repeat the estimation process but now use a less noisy sensor whose output y_k is given by:

$$y_k = \tilde{y}_k + 0.001 * \text{randn}(N, 1)$$

Does the value of x change appreciably?

- f) Comment on how you would validate the accuracy of the estimate you generated.

¹As of the moment I was writing this homework set (September 5, 2020), the Wikipedia Article on ARMA is rather verbose and hard to follow. You can read it to learn what an ARMA model is but I would suggest looking elsewhere to learn about ARMA.