ENEE 691 Homework 3.v0

Spring 2023

Please submit your homework as a zip file which should include a pdf file explaining your solutions and your codes/notebook. Due date: 9:00 PM, March 31, 2023.

Introduction

For this homework, you will be working on a differential reflectance dataset. Figure shows a typical reflectance measurement setup. The one on the left measures reflectance from a p-polarized laser, the one on right from an s-polarized laser. [1] The wavelength of the laser is 400 nm. We place a 0.7 nm thick film on a substrate whose complex refractive index is $\tilde{n} = n + ik$, where the real part n is the refractive index and indicates the phase velocity, while the imaginary part k is called the optical extinction coefficient or absorption coefficient. We work either with SiO_2-coated silicon substrates or glass substrates. We change the incidence angle (θ_{inc}) from 0 degrees to 60 degrees at a step of 1 degrees and we record the intensity of the reflected light with the help of a photodiode. When we use reflectance from the bare substrate as a reference, then we can get the differential reflectance by simply calculating $(\Gamma - \Gamma_0)/\Gamma_0$, where Γ and Γ_0 are the reflectances from the sample coated substrate and bare substrate.

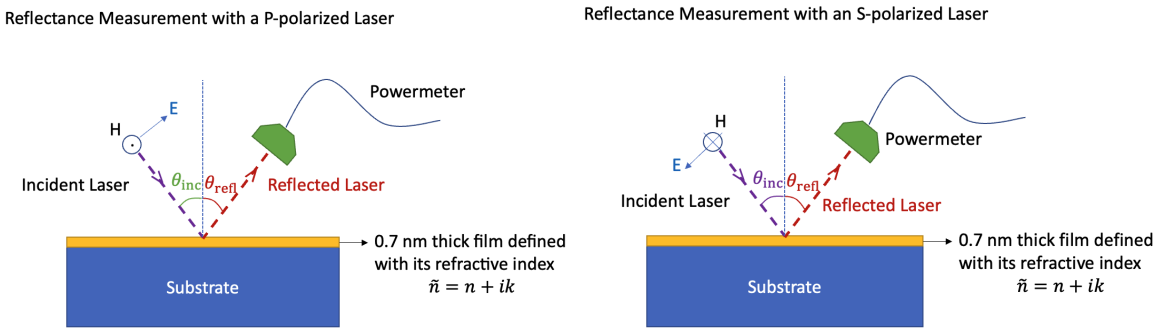

Figure 1: Reflectance measurement setup for (left) p-polarized and (right) s-polarized light.

Please the three datasets (csv) files from https://github.com/simsekergun/ENEE691/tree/main/homework3 or use the cell below to get a copy of them in your jupyter notebook.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

X_train = pd.read_csv('https://raw.githubusercontent.com/simsekergun/ENEE691/main/homework3/Reflectance_X_train.csv')
Y_train = pd.read_csv('https://raw.githubusercontent.com/simsekergun/ENEE691/main/homework3/Reflectance_Y_train.csv')
X_test = pd.read_csv('https://raw.githubusercontent.com/simsekergun/ENEE691/main/homework3/Reflectance_X_test.csv')
n_train = Y_train.iloc[:,1]
k_train = Y_train.iloc[:,2]
```

"Reflectance_X_train.csv" has 244 columns as follows.

- (a) Columns 1-61: reflectance data as a function of incidence angle (from 0 degrees to 60 degrees) of p-polarized 400 nm laser from a material with a thickness of 0.7 nm placed on top of a SiO_2/silicon substrate,
- (b) Columns 62-122: similar to (a) for s-polarized laser,

[1] These letters have a German origin: s = senkrecht = perpendicular, p = parallel. It basically tells whether the electric field is perpendicular or parallel to the plane of incidence.

- (c) Columns 123-183: similar to (a) for a glass substrate,
- (d) Columns 184:244: similar to (c) for s-polarized laser.

Each row corresponds to a unique (n, k)-pair given in "Reflectance_Y_train.csv". As you can see, n changes between 0.1 and 6 and k changes between 0 and 5, both at the steps of 0.1.

Your task is finding the (n, k) values for the reflectance data given in Reflectance_X_test.csv, which is in the same format as the Reflectance_X_train.csv.

TASK-1 (30 points)
Use linear regression and predict the complex refractive index, in other words n and k values for the unknown material. The reflectance from this unknown material is the "Reflectance_X_test.csv". Do not forget to report your training accuracy both for predicting n and k.

TASK-2 (40 points)
Note that we have 60 unique values of n and 51 unique values of k. Hence this problem (finding n and k values) can be done with multinomial logistic regression as well. You need to treat each unique value of n (first, then of k) as a class.
2.1 Build a multinomial logistic regression model to predict the most probable four n-values this unknown material that might have. You will need to use a lot of iterations (e.g., set max_iter = 100000). Create a table something like this
Then use the following formula to get your final prediction for n_f

Table 1: The most probable four n values predicted with the logistic regression and their probabilities.

Guess No.	Prediction n value	Probability
1	n_1	p_1
2	n_2	p_2
3	n_3	p_3
4	n_4	p_4

$$n_f = \frac{\sum_{i=1}^{4} n_i p_i}{\sum_{i=1}^{4} p_i} \tag{1}$$

Use linear regression and predict the complex refractive index, in other words n and k values for the unknown material. The reflectance from this unknown material is the "Reflectance_X_test.csv".
2.2 Repeat task **2.1** to estimate the k.
2.3 Compare the n and k values you determined at the end of Task 1 and 2.

TASK-3 (40 points)
TASK-3.1 Use SVMs to predict n and k. Since I haven't done this step by myself, I have no clue what kind of kernel we should use. Try different approaches and come up with your best solution.
TASK-3.2 Compare the performances of regression via classification done with logistic regression vs. SVMs.
TASK-3.3 The true (experimentally measured) refractive index of this material is 2.5 + 3.74i. Calculate the error for each solution, i.e.,
Linear Regression Error (real part): ...%
Linear Regression Error (imag. part): ...%
Which one is the most accurate?