M4S18A2 Machine Learning - Coursework 2

Omar Haque

March 15, 2018

Introduction

This report has 3 sections. In the first, I implement a K-means clustering algorithm, and a Gaussian Mixture Model under the Expectation-Maximisation framework. I use these algorithms to create a program which automatically segments and counts the number of cells in a medical image. In the second part I explore the role of bias in modern AI applications. In the third section, I implement a Hidden-Markov Model in order to create a predictive trading model for Ethereum prices.

Question 1

Here is the implementation for the K-means algorithm

```
def k_means_clustering(X_cleaned,K=10,maxiter=200):

    # get dimensions of data
    n = X_cleaned.shape[0]
    dim = X_cleaned.shape[1]

    ## initialise centroids randomly
    random.seed(1)
    centroids = X_cleaned[random.sample(range(n),K),:]

    # initialise convergence criteria
    converged = False

    for i in range(maxiter):

        # bookkeeping, simply keep track of the old centroids
        old_centroids = centroids

        # data assignment step
        data_clustering = get_closest_cluster(X_cleaned, centroids,K)
        # centroids update step
```

```
        centroids = update_centroids(X_cleaned, data_clustering,K)

        if (np.allclose(centroids,old_centroids,rtol=1e-3)):
            print("converged at iteration " + str((i+1)))
            converged = True
            break

    if (not converged):
        print("Warning, clusters did not converge.")

    return centroids, data_clustering
```

1 Theory

Here give a brief summary of the physical effect of interest and provide necessary equations. Here is how you insert an equation. According to references [1–3] the dependence of interest is given by

$$u(\lambda, T) = \frac{8\pi hc\lambda^{-5}}{e^{hc/\lambda kT} - 1}, \tag{1}$$

where T is temperature in Kelvin, c is the speed of light, etc. Don't forget to explain what each variable means the first time that you introduce it.

2 Procedures

Give a schematic of the experimental setup(s) used in the experiment (see figure 1). Give the description of abbreviations either in the figure caption or in the text. Write a description of what is going on.

Don't forget to list all important steps in your experimental procedure!

Use active voice either in past or present through all the report and be consistent with it: The laser light comes from to ... and eventually arrived to the balanced photodiode as seen in the figure 1.

Sentences in the past voice while correct are generally considered hard to read in large numbers. The laser light was directed to ..., wave plates were set to ... etc.

3 Analysis

In this section you will need to show your experimental results. Use tables and graphs when it is possible. Table 1 is an example.

Analysis of equation 1 shows ...

Note: this section can be integrated with the previous one as long as you address the issue. Here explain how you determine uncertainties for different measured values. Suppose that in the experiment you make a series of measurements of a resistance of the wire R for

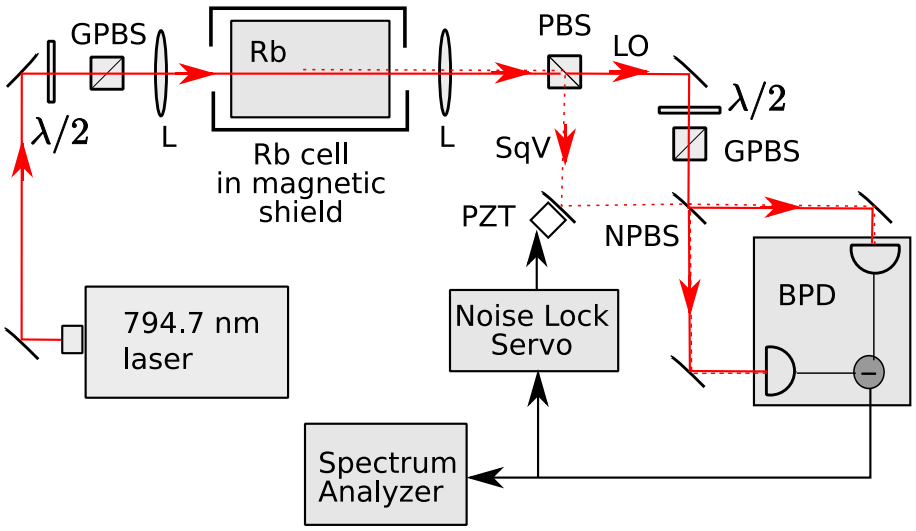

Figure 1: Every figure MUST have a caption.

Table 1: Every table needs a caption.

x (m)	V (V)
0.0044151	0.0030871
0.0021633	0.0021343
0.0003600	0.0018642
0.0023831	0.0013287

different applied voltages V, then you calculate the temperature from the resistance using a known equation and make a plot temperature vs. voltage squared. Again suppose that this dependence is expected to be linear [2], and the proportionality coefficient is extracted from the graph. Then what you need to explain is that for the resistance and the voltage the uncertainties are instrumental (since each measurements in done only once), and they are .... Then give an equation for calculating the uncertainty of the temperature from the resistance uncertainty. Finally explain how the uncertainty of the slop of the graph was found (computer fitting, graphical method, *etc.*)

If in the process of data analysis you found any noticeable systematic error(s), you have to explain them in this section of the report.

It is also recommended to plot the data graphically to efficiently illustrate any points of discussion. For example, it is easy to conclude that the experiment and theory match each other rather well if you look at Fig. 1 and Fig. 2.

4 Conclusions

Here you briefly summarize your findings.

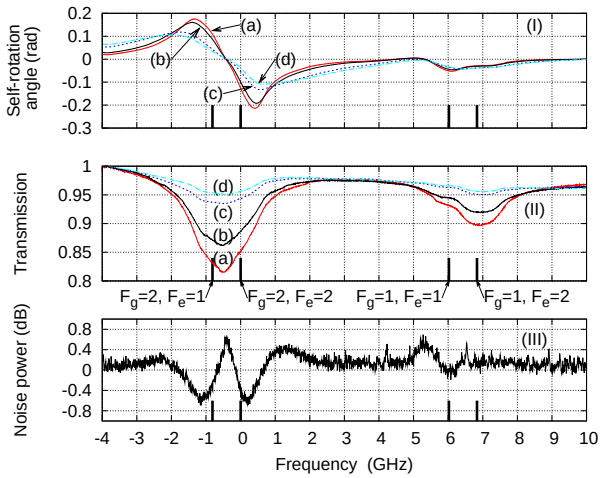

Figure 2: Every plot must have axes labeled.

References

[1] A. C. Melissinos and J. Napolitano, *Experiments in Modern Physics*, (Academic Press, New York, 2003).

[2] N. Cyr, M. Têtu, and M. Breton, IEEE Trans. Instrum. Meas. **42**, 640 (1993).

[3] *Expected value*, available at `http://en.wikipedia.org/wiki/Expected_value`.