

CS 598 PS  
Machine Learning for Signal Processing  
Problem Set 1

Christian Howard  
howard28@illinois.edu

**Abstract**

Within this report, problems spanning probability, tensors, and signal processing were tackled and solved. To grab some extra credit, two sounds were recorded and were used to produce a spectrogram using the code created in the last problem.

## Contents

<b>1</b>	<b>Problem 1 - A Probability Problem</b>	<b>3</b>
<b>2</b>	<b>Problem 2 - Manipulating Data using Linear Algebra</b>	<b>4</b>
2.1	. . . . .	4
2.1.a	. . . . .	4
2.1.b	. . . . .	4
2.2	. . . . .	5
2.2.a	. . . . .	5
2.2.b	. . . . .	5
<b>3</b>	<b>Problem 3 - Signal Processing is Linear Algebra</b>	<b>6</b>

## 1 Problem 1 - A Probability Problem

For this first problem, the goal is to find the probability a woman has breast cancer given she has a positive mammograph. To present the solution to this problem, let me first define a few things that will prove useful. First, below is the notation for the various events:

$$\begin{aligned} R_1 &= \text{woman has breast cancer} & R_2 &= \text{woman does not have breast cancer} \\ M^+ &= \text{mammogram is positive} & M^- &= \text{mammogram is negative} \end{aligned}$$

In the problem statement, we are given the following information:

- The probability that a woman has breast cancer is 0.8%
- If a woman has breast cancer, there is a probability of 90% that the mammogram will be positive
- If she has no breast cancer, the probability of a positive mammogram is 7%

Based on the above information, we can find the following probabilities:

$$\begin{aligned} P(M^+|R_1) &= 0.9 \\ P(M^+|R_2) &= 0.07 \\ P(R_1) &= 0.008 \\ P(R_2) &= 0.992 \end{aligned}$$

Given the above information and definitions, our goal is to find  $P(R_1|M^+)$ . Using Baye's Rule, this becomes straight forward. The following is how we can arrive to the desired probability:

$$\begin{aligned} P(R_1|M^+) &= \frac{P(M^+|R_1)P(R_1)}{P(M^+)} \\ &= \frac{P(M^+|R_1)P(R_1)}{P(M^+|R_1)P(R_1) + P(M^+|R_2)P(R_2)} \\ &= \frac{(0.9)(0.008)}{(0.9)(0.008) + (0.07)(0.992)} \\ &= 0.0939 \\ &= 9.39\% \end{aligned}$$

Thus we find that the probability of a woman having breast cancer, given they got a positive mammogram, is 9.39%.

## 2 Problem 2 - Manipulating Data using Linear Algebra

### 2.1

I am given  $K$  grayscale images such that each image has dimension  $M \times N$ . For use as a data set, each image is vectorized and stuffed into a large data set matrix,  $D$ , that has dimension  $(MN) \times K$ .

#### 2.1.a

The goal for this part is to return the mean image, let us call it  $\mu_I$ , for the above data set and return it as an  $M \times N$  matrix.

Given the data set is represented by the matrix  $D$ , one can multiply  $D$  by a vector that will do the averaging for the vectorized form of the images and then invert the vectorization of the result. This final operation can be defined as the following:

$$\mu_I = (D\mathbf{v})^{(M)}$$

where  $\mathbf{v} = [v_1, v_2, \dots, v_n]^T$  is defined such that  $v_i = \frac{1}{K} \forall i$  and  $n = MN$ .

#### 2.1.b

The goal for this part is to compute a  $2 \times 2$  covariance matrix where this covariance is based on the mean color for the top and bottom half of a given image for all  $K$  images in the data set.

Based on looking at patterns with simple  $D$  values, a transformation was found for going from a vectorized image to a two element vector with the two half image averages we desire. For notational simplicity, let us define  $I_p$  as an identity matrix of size  $p \times p$ . Using this definition, the transformation matrix can be defined like so:

$$H_{(M,N)} = \frac{1}{N} \mathbf{1}_{(N)}^T \otimes I_2 \otimes \left( \frac{2}{M} \mathbf{1}_{(M/2)}^T \right)$$

where  $\mathbf{1}_{(p)} \in \{1\}^{p \times 1}$  (i.e. a column vector of 1s of dimension  $p$ ). Using this matrix, and using the same  $\mathbf{v}$  from the last problem, we can compute the covariance  $\Sigma$  using the following:

$$\begin{aligned} D_\mu &= H_{(M,N)} D \\ \boldsymbol{\mu} &= D_\mu \mathbf{v} \\ \boldsymbol{\delta} &= D_\mu - \mathbf{1}_{(K)}^T \otimes \boldsymbol{\mu} \\ \Sigma &= \frac{1}{K-1} \boldsymbol{\delta} \boldsymbol{\delta}^T \end{aligned}$$

## 2.2

To make things more interesting, now we are given a set of  $K$  colored images where this whole data set is represented as a  $M \times N \times 3 \times K$  tensor called  $D$ , where the dimension of size 3 is the color channel dimension.

### 2.2.a

For this part we are asked to compute the average image across both the  $K$  data samples and 3 color channels, which I will define as  $\mu_I$ . First, for notational simplicity, let us define  $I_p$  as an identity matrix of size  $p \times p$ . With this, we can find an operator to find the average image in its vectorized form using the following operation:

$$\mu_I = \left( \left( \frac{1}{K} \mathbf{1}_{(K)}^T \otimes \frac{1}{3} \mathbf{1}_{(3)}^T \otimes I_N \otimes I_M \right) \text{vec}(D) \right)^{(M)}$$

Working left to right in the Kronecker product sequence, the first term represents an averaging row vector over the  $K$  pieces of data. The second term from the left then specifies an averaging row vector over the 3 color channels. This vector then undergoes a Kronecker product with a sequence of identity matrices of dimensions  $N \times N$ , and  $M \times M$  going from left to right. The resulting matrix, after all the Kronecker products, will take the vectorized data set and produce an  $M \times N$  average image  $\mu_I$ .

### 2.2.b

This next part is to instead compute the average image within just the red channel of the matrix. We will keep in mind that since the image is RGB, red will be the first channel in the third tensor dimension. With this defined, we can find the average  $M \times N$  red channel image,  $\mu_r$ , using the below operation:

$$\mu_r = \left( \left( \frac{1}{K} \mathbf{1}_{(K)}^T \otimes [1, 0, 0] \otimes I_N \otimes I_M \right) \text{vec}(D) \right)^{(M)}$$

Using the above operation, the second term  $[1, 0, 0]$  effectively zeros out contributions from the green and blue channels, producing an average vectorized  $M \times N$  image which represents the average red channel image. We then vec-transpose the result to obtain an  $M \times N$  average red image.

### 3 Problem 3 - Signal Processing is Linear Algebra

Within this problem, the goal is to find an expression for a matrix  $A$  such that one can multiply this matrix by some input sound  $\mathbf{x}$  and produce the vectorized complex-valued spectrogram for this sound. Some desired traits for this matrix  $A$  are the following:

- The transformation will be based on a 1024 point Discrete Fourier Transform (DFT)
- The transformation will use a Hann window
- The transformation will be based on a hop size of 512 points

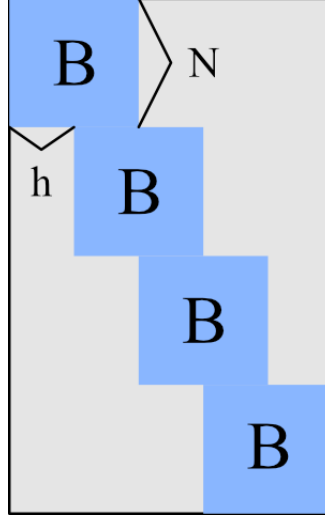
For notational convenience, let us define  $N$  as the number of points having the windowing and DFT performed on them. Let us also define  $h$  as the hop size value being used and  $M$  as the number of windows we will be processing. To formulate this matrix  $A$ , let us first take a look at a local window of  $N$  data points. The transformation needed to process this data is the matrix  $B$ , defined below as:

$$B = DH$$

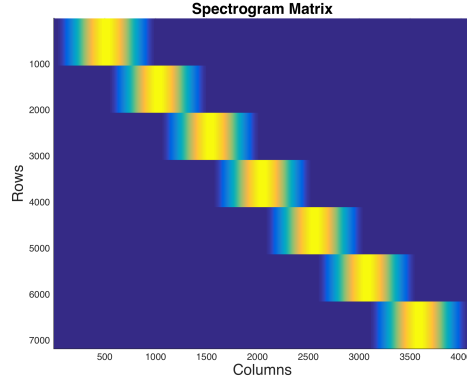
where  $D \in \mathbb{C}^{N \times N}$  is the Fourier Matrix defined as  $D_{jk} = \frac{1}{\sqrt{N}} e^{2\pi i jk/N}$  and  $H = \text{diag}(\mathbf{w})$  represents the Hann window operator, where  $\mathbf{w} = [w_0, \dots, w_{N-1}]^T$  and the elements of  $\mathbf{w}$  can be defined as:

$$w_k = \frac{1}{2} \left( 1 - \cos \left( \frac{2\pi k}{N-1} \right) \right)$$

Given the definition of matrix  $B$ , we can then construct the overall matrix  $A$  by embedding  $B$  within  $A$  as  $M$  different submatrices shifted  $h$  columns to the right and  $N$  rows down relative to where the previous submatrix is located. This can be visualized by using the structure shown below:



Using this structure, an example graphic of  $\|A\|$  with the DFT and Hann window built into it is the following:



Now with matrix  $A$  defined properly, we can then produce a  $N \times M$  spectrogram matrix  $S$  using the following:

$$S = (A\mathbf{x})^{(N)}$$

Using this formulation for  $A$  and the specific constraints mentioned above, i.e.  $N = 1024$ ,  $h = 512$ , below are two sample spectrograms based on two sounds that I recorded so I can grab some of that extra credit:

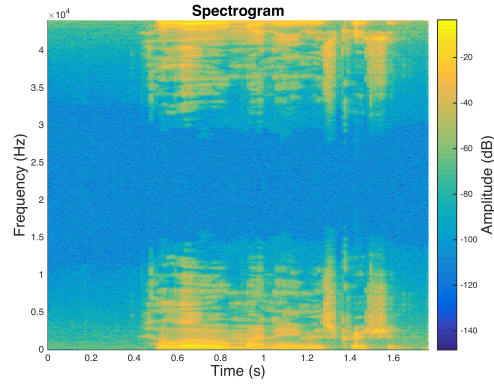


Figure 1: Spectrogram from saying "Hello my name is Christian"

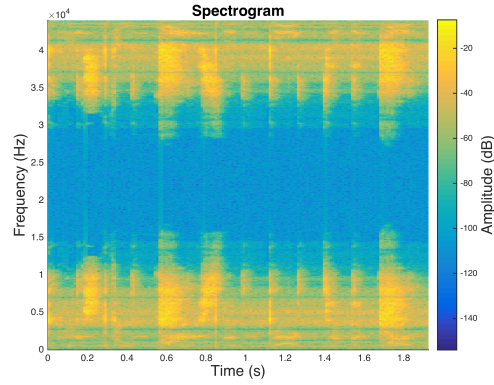


Figure 2: Spectrogram from random 2 second portion of Hollywood Undead song