

Homework 4: Bacteria Dataset Preprocessing

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN

DUE 11/05/2018

This homework is intended as a milestone of the course project. I suggest that you use Julia, but you may use Matlab, Python, or your preferred coding language. The dataset can be found at <https://www.dropbox.com/sh/11p3ke597r9f1dt/AADh0tyEVroUKHd9Zw8XRUZxa?dl=0>.

- (a) Load a randomly selected image \mathbf{X}_0 from the dataset.
- (b) Transform \mathbf{X}_0 to grayscale, to obtain \mathbf{X}_{gray} .
- (c) Rescale \mathbf{X}_{gray} to obtain a 28×28 image \mathbf{X}_{28} .
- (d) Vectorize \mathbf{X}_{28} to obtain a vector $\mathbf{x}_i \in \mathbb{R}^{784}$.
- (e) From the Excel file containing the information of interest in the project, load the row corresponding to the image you selected, and store it in a vector \mathbf{y}_i .
- (f) What is the size of \mathbf{y}_i ?

Deliver your code for each of the points above.