

Assignment 3

1. Load the file `WineData.mat` containing a matrix $\mathbf{X} \in \mathbb{R}^{13 \times 178}$ of chemical analysis data of wines derived from the same area of Italy but originating from three different cultivars. The attributes in each column of \mathbf{X} are

- 1 Alcohol
- 2 Malic acid
- 3 Ash
- 4 Alcalinity of ash
- 5 Magnesium
- 6 Total phenols
- 7 Flavanoids
- 8 Nonflavanoid phenols
- 9 Proanthocyanins
- 10 Color intensity
- 11 Hue
- 12 OD280/OD315 of diluted wines
- 13 Proline

The vector \mathbf{I} is the annotation vector indicating from which one of the three cultivars the wine originates. Using the division of the data into three clusters defined by the annotation vector, find the first three LDA directions, and plot the three different two-dimensional projections of the wine data. Comment on how well the clusters are identifiable in this plot.

2. Using again the handwritten digits data, select two different digits, for instance '0' and '4'. Denote by \mathbf{X}_{04} the data matrix containing the images.
 - (a) Plot the data projected on the first two LDA directions.
 - (b) Write your ANLS algorithm to find the rank- k non-negative factorization

$$\mathbf{X}_{04} \approx \mathbf{W}\mathbf{H}.$$

Run the algorithm with different values of k , e.g., $k = 5, 10, 20$, and plot the feature vectors as images. You should see pieces of the digits '0' and '4' in those images. You may use the built-in Matlab function `lsqnonneg` in your algorithm.

Pick one of the feature vectors that clearly looks like a piece of a '0', and verify that the coefficients in the matrix \mathbf{H} that correspond to that feature vector reveal if the corresponding data vector represents a '0' or a '4'.