

Worksheet 2

Conor Houghton

1 Introduction

There are two main tasks in this worksheet; the first is fairly simple, take the matrix

$$A = \begin{pmatrix} 6 & -2 \\ -2 & 9 \end{pmatrix} \quad (1)$$

and work out its eigenvectors and eigenvalues, write it in the form PDP^T and check by multiplying that out that it gives back the origin matrix.

The second is a bit harder, in the github¹ you will find a file called `ws2_india.csv`. This is a table of how many people out of 10,000 speak each of the scheduled languages of India in each of India's regions. The concept of scheduled languages is a bit slippery, these are in some sense the 'main languages' but include Sanskrit which has almost no speakers, in addition, some other languages have been bundled up with the scheduled languages as if they were dialects rather than languages while other languages are listed as 'non-scheduled' languages. It is all a bit of a mess, but either way, we are just interested in these data as example data and it is impressive these data are available at all!

The idea here is to perform PCA on these data, regarding each region as a data point. Find the eigenvalues, plot the first two principal components and see if you can interpret them. The interpretation will be, admittedly, easier for some of you than others. Experiment with using the covariance matrix or the correlation matrix.

¹github.com/emadm0067/2025_26/tree/master/worksheets