Rensselaer

why not change the world?®

Lab exercises: Principal Component Analysis

Ahmed Eleish

ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960 Lab 4, November 8th, 2024

Tetherless World Constellation
Rensselaer Polytechnic Institute

Lab 04

https://archive.ics.uci.edu/dataset/109/wine

Principal Component Analysis

Using the wine dataset:

- Compute the PCs and plot the dataset using the 1st and 2nd PC.
- Identify the variables that contribute the most to the 1st PC.
- Train a classifier model to predict wine type using the 13 attributes.
- Train a classifier model to predict wine type using the data projected into the first 3 PCs.
- Drop the variables least contributing to the 1st PC and rerun PCA.
- Train a classifier model to predict wine type using the data projected into the first 3 PCs after rerunning PCA.
- Compare the 3 classification models using contingency tables and prevision/recall/f1 metrics

Please push to your github repository:

1. All your code in a *.R or *.MD file
2. All text outputs (contingency tables)
3. All plots (group colored scatter plots, kNN accuracy plots, k-Means “elbow” plots)

Thanks!
Have a great weekend!