

# **STA 6707: HOMEWORK 1**

**Due in class Thu Feb 4**

## **Problem 1:**

The data set USTATES-data.txt (available in the folder Data Sets) contains the unemployment rates of the 50 US States, plus those of the District of Columbia and Puerto Rico for the period Jan 1976- March 2007.

Use PCA analysis to summarize the basic patterns in the data and carefully comment on the results.

Specifically, discuss in your report possible transformations of the variables, exclusion of outliers, choice of number of principal components, fit of the solution, interpretation of the principal components.

Total length: 3 pages

## **Problem 2:**

This data set contains information about the chemical composition of wines. The variables are described next:

- (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

Perform PCA and comment on the results.

Total length: 2 pages

## Homework Preparation Guidelines

Your reports should be well-organized, clear and to the point. More specifically:

- (1) Describe the particular question being addressed.
- (2) Introduce the variables that you are going to use (this includes describing any transformations you think appropriate)
- (3) Describe the data you are going to use (variables and objects, including transformations of variables and/or removal of objects/variables you think appropriate and explain why)
- (4) Describe the results (preferably in tabular form) (NO COMPUTER OUTPUT)
- (5) Summarize the results and provide comments
- (6) If necessary include graphs and refer to them in your text.
- (7) Organize your report in the appropriate sections (Data processing issues, results, comments)
- (8) Start your report with a short executive summary.