Programming questions

• **USArrests data set**
  • `d <- USArrests`
  • Print a vector of the state names from the highest Assault rate to the lowest Assault rate.
  • Produce a scatter plot of Rape and Murder.
  • Compute Pearson correlation coefficient for Murder and Rape, Rape and UrbanPop, Murder and UrbanPop.
  • Run K-Means clustering algorithm for K=3 and experiment with Assault and Rape
  • Run K-Means clustering algorithm for K=3 and experiment with all the four features.
  • Use the Elbow Method to find an optimal value of K

• **(Hierarchichal) clustering**
  • Feature scaling
  • NLI data set (75 documents, 5 languages)
Programming questions

- **Auto data set**
  - Compute Pearson's correlation coefficients for mpg, displacement, weight, horsepower, acceleration in the Auto data set
  - Draw boxplots to visualize comparison mpg by origin, mpg by model year, and weight by origin
Feature scaling

Different ranges and units of features

• Is the engine displacement more significant than mpg/cylinders/acceleration?

> str(Auto)
'data.frame': 392 obs. of 9 variables:
$ mpg : num 18 15 18 16 17 15 14 14 14 15 ...
$ cylinders : num 8 8 8 8 8 8 8 8 8 8 ...
$ displacement: num 307 350 318 304 302 429 454 440 455 390 ...
$ horsepower : num 130 165 150 150 140 198 220 215 225 190 ...
$ weight : num 3504 3693 3436 3433 3449 ...
$ acceleration: num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
$ year : num 70 70 70 70 70 70 70 70 70 ...
$ origin : Factor w/ 3 levels "USA","Europe",...: 1 1 1 1 1 1 1 1 1 ...
$ name : Factor w/ 304 levels "amc ambassador brougham",...: 49 36 231
Feature scaling

Scaling

- normalization \( z = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \), i.e., the feature values are shifted and rescaled so that they end up ranging between 0 and 1, \( z \in \langle 0, 1 \rangle \)

- standardization \( z = \frac{x - \bar{x}}{sd_x} \), i.e., the feature values are centered around the mean with a unit standard deviation \( \bar{z} = 0, sd_z = 1 \)

Useful especially for
- Gradient Descent Based Algorithms
- Distance based algorithms
Tobespecializedinone specific subject is also a good thing. It will make you a professional in a specific subject, but when you it comes to society they will consider you own and it also money.

Changing your habits is not an easy thing, but one is urged to do it for a number of reasons. A successful teacher should renew his lectures so they will be considered an invention in its own, and it also money.

It is troublesome, as it seems, but it keeps me fresh in information with a good status among my colleagues.

Alternating work and dormancy in your life pace with activities and exercise of great benefit to the body and the mind.

Lastly, change is a difficult decision in the human, but it is important for many good reasons, and gets back on the human with good benefits.

This indicates the to date and also his mind constant. Also, this reputation colleagues.

You are constantly trying to change your ways of making, or synthesizing chemical compounds, you might come out with less expensive methods.

When I change my glasses color, for example, this would be attractive to my students and colleagues, and make me feel better.

In contrast, many people believe that changing is really important in people’s lives.

Changing your habits is not an easy thing, but one is urged to do it for a number of reasons. A successful teacher should renew his lectures so they will be considered an invention in its own, and it also money.

It is troublesome, as it seems, but it keeps me fresh in information with a good status among my colleagues.

Alternating work and dormancy in your life pace with activities and exercise of great benefit to the body and the mind.

Lastly, change is a difficult decision in the human, but it is important for many good reasons, and gets back on the human with good benefits.

This indicates the to date and also his mind constant. Also, this reputation colleagues.

You are constantly trying to change your ways of making, or synthesizing chemical compounds, you might come out with less expensive methods.

When I change my glasses color, for example, this would be attractive to my students and colleagues, and make me feel better.

In contrast, many people believe that changing is really important in people’s lives.
Identifying the native language (L1) of a writer based on a sample of their writing in a second language (L2)

Our data

- **L1s**: Arabic (ARA), Chinese (ZHO), French (FRA), German (DEU) Hindi (HIN), Italian (ITA), Japanese (JPN), Korean (KOR), Spanish (SPA), Telugu (TEL), Turkish (TUR)
- **L2**: English
- **Real-world objects**: For each L1, 1,000 texts in L2 from The ETS Corpus of Non-Native Written English (former TOEFL11), i.e. *Train ∪ DevTest*
- **Target class**: L1

More detailed info is available at the course website.
**Topic**
Most advertisements make products seem much better than they really are

**Sample text**
now a days the publisity is the best way to promoved a produt and if you wanth to sale a product you should bring some information that makes , that the people who is seeing the advertisements make sure that the product very good and in the future this person could buy it .

*L1 = Spanish*