Programming questions

• (Hierarchical) clustering
  • Feature scaling
  • NLI data set (75 documents, 5 languages)

• Gradient descent algorithm
  • Find a minimum of a function using Gradient Descent Algorithm (simple illustration)

• 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

• Linear regression
  • Auto data set, target attribute: mpg
Feature scaling

Different ranges and units of features

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

```r
> str(Auto)
'data.frame': 392 obs. of 9 variables:
$ mpg : num 18 15 18 16 17 15 14 14 15 ...
$ cylinders : num 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 1 ...
$ name : Factor w/ 304 levels "amc ambassador brougham",...: 49 36 231
```
Feature scaling

Scaling

• normalization $z = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$
  $z \in < 0, 1 >$

• standardization $z = \frac{x - \bar{x}}{sd_x}$
  $\bar{z} = 0$, $sd_z = 1$

Useful especially for

• Gradient Descent Based Algorithms
• Distance based algorithms
This indicates that to date and also his mind constant or static, you are consistent. 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.

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

The newer method is considered an invention in its own, and it also saves money.

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.

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

Also, you might think of changing your view, clothes or even your hair cut.

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

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

This indicates that to date and also his mind constant or static, you are consistent.
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. $Train \cup DevTest$
- **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**
Linear regression

Random error term

- numerical target attribute $Y$
- $y = X\Theta^T + \epsilon$
- random error term $\epsilon$ having mean zero, very often unobserved
Linear regression
Random error term

- $\epsilon_i = y_i - \Theta^\top x_i$ (true target value $y_i$, expected value $\Theta^\top x_i$)
- Assumption like: At each value of $A_1$, the output value $y$ is subject to random error $\epsilon$ that is normally distributed $N(0, \sigma^2)$
Linear regression
Random error term

- $\epsilon_i = y_i - \Theta^\top x_i$ (true target value $y_i$, expected value $\Theta^\top x_i$)
- residual $e_i = y_i - \hat{\Theta}^\top x_i$ is an estimate of $\epsilon_i$