Introduction to Machine Learning
NPFL 054
http://ufal.mff.cuni.cz/course/npfl054

Barbora Hladká        Martin Holub

{Hladka | Holub}@ufal.mff.cuni.cz

Charles University,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
**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 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 \( z = \frac{x-x_{\text{min}}}{x_{\text{max}}-x_{\text{min}}} \), 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 \( z = \frac{x-x}{sdx} \), 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

```r
> head(scale(Auto[,c('mpg', 'displacement', 'weight')])))
   mpg      displacement      weight
1 1.075915  0.6197483  1.075915
2 1.486832  0.8422577  1.486832
3 1.181033  0.5396921  1.181033
4 1.047246  0.5361602  1.047246
5 1.028134  0.5549969  1.028134
6 2.241772  1.6051468  2.241772
```
Tobespecializedinone
specificsubjectisalsoa
goodthingitwillmakeyoua
profissionalinaspecific
subject,butwhenyouit
comes tosocietytheywill

Proficiency: high
Prompt: P2

Changing yourhabbitsis not
an easy thing, but one is
urged to do it for a number
of reasons.
A successful teacher should
renew his lectures so

Proficiency: high
Prompt: P2

Thisindicates that
todate and also
his mind constant.
Also, this
reputation
colleagues.
Proficiency: low
Prompt: P1

To be specialized in one
specific subject is also a
good thing it will make you a
profissional in a specific
subject, but when you it
comes to society they will

Proficiency: high
Prompt: P2

The newer method is
considered an own, and it also
profit money.

Proficiency: medium
Prompt: P2

In contrast, many people
believe that changing is
truly important in people's
lives.

Proficiency: high
Prompt: P2

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

Proficiency: high
Prompt: P2

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

Proficiency: low
Prompt: P7

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

Proficiency: medium
Prompt: P2

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

Proficiency: low
Prompt: P7

Changing yourhabbitsis not
an easy thing, but one is
urged to do it for a number
of reasons.
A successful teacher should
renew his lectures so

Proficiency: high
Prompt: P2

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

Proficiency: high
Prompt: P2

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

Proficiency: low
Prompt: P7

University, I keep getting
newer information from
different books and
resources and include it in
my lectures.

Proficiency: low
Prompt: P8

Thisindicates that
todate and also
his mind constant.
Also, this
reputation
colleagues.
Proficiency: low
Prompt: P1

You can consider
changing your ways
making, or synthesizing,
chemical compounds, you
might come out with less
expensive methods.

Proficiency: medium
Prompt: P2
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. $\text{Train} \cup \text{DevTest}$
- **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
Nowadays, publicity is the best way to promote a product, and if you want to sell a product, you should bring some information that makes sure that the people who are seeing the advertisements make sure that the product is very good and in the future, this person could buy it.

L1 = Spanish
• numerical target attribute $Y$
• $y = X\Theta^T + \epsilon$
• random error term $\epsilon$ having mean zero, very often unobserved
Linear regression
Random error term

- $\epsilon_i = y_i - \Theta^T x_i$ (true target value $y_i$, expected value $\Theta^T x_i$)
- Assumption like: At each value of $A_1$, the output value $y$ is subject to random error $\epsilon$ that is normally distributed $N(0, \sigma^2)$
• $\epsilon_i = y_i - \Theta^T x_i$ (true target value $y_i$, expected value $\Theta^T x_i$)
• residual $e_i = y_i - \hat{\Theta}^T x_i$ is an estimate of $\epsilon_i$