machine learning

Rita Ribeiro (coordinator)
Carlos Soares (coordinator)
João Mendes Moreira
Nuno Moniz

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

U.PORTO
FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

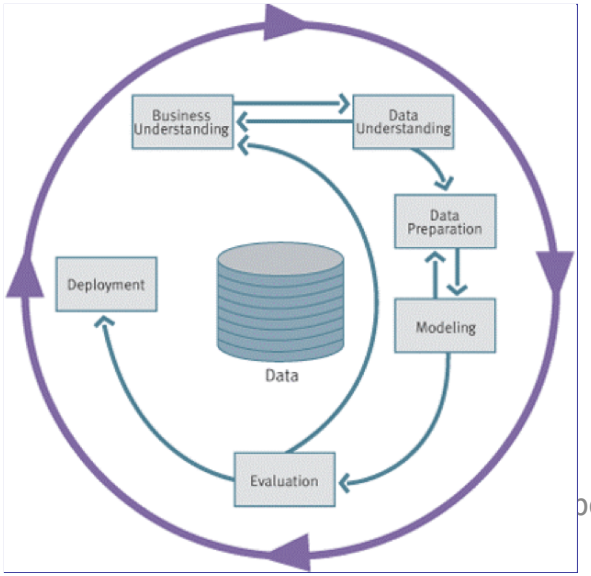
UNIVERSIDADE DO PORTO

introduction to the course

- assumptions
- goals
- plan
- evaluation
- team
- support
- software
- bibliography

we assume superficial knowledge on (from AI course)

- develop simple predictive data mining projects involving the most traditional tasks
  - classification
  - regression



- namely
  - identify problems that can be addressed with predictive DM
  - follow proper methodology to solve DM problems
    - simple
  - evaluate results
    - technical perspective
    - ... and application domain

UNIVERSIDADE DO PORTO

so, our goals are

(well, your goals, really…)

- consolidate/complete basic skills in DM
  - predictive tasks
  - descriptive tasks
  - project development
- acquire (some) advanced skills in ML/DM
  - including on research topics of the lecturers

UNIVERSIDADE DO PORTO

plan

week	date T lesson	T lecturer	T session	TP session
1	21/10/2021	CS	Introduction to course. Data mining methodologies. (DM project) introduction	-
2	28/10/2021	CS	Classification: introduction, decision trees & evaluation. Scoring with classification model: approach & evaluation	(DM project) suggested goal: dataset(s) based on a single table; 1st kaggle submission;
3	4/11/2021	RR	DM projects: data preparation. Data preparation methods (basic). Data preparation in R	(DM project) suggested goal: datasets based on multiple tables with engineered features
4	11/11/2021	RR	DM projects: data understanding (quality & visualization). Data visualization in R	(DM project) suggested goal: data understanding
5	18/11/2021	CS	Advanced issues in data preparation and modeling (feature selection, dimensionality reduction, imbalanced class distribution)	(DM project) suggested goal: application of methods for class imbalance
6	25/11/2021	CS	Regression. Learning algorithms and ensembles	(DM project) suggested goal: final submission; analysis of models obtained with different algorithms
7	2/12/2021	RR	Clustering	(DM project) suggested goal: customer profiling
8	9/12/2021	RR	Frequent pattern mining	
9	16/12/2021	CS	Recommender Systems	(DM project) close project
-	23/12/2021			
-	30/12/2021			
10	6/1/2022	RR	Anomaly detection	hands-on (FPM + RS)
11	13/1/2022	RR	Neural Networks & Deep Learning	hands-on (AD)
12	20/1/2022	CS	AutoML & Metalearning	hands-on (NN & DeepL)
13	27/1/2022	CS	Seminar	hands-on (autoML & metalearning)

evaluation: test

- model available in moodle
  - sometime soon...
- grade
  - 50% of final grade
  - minimum grade of 7/20

evaluation: competition (1/2)

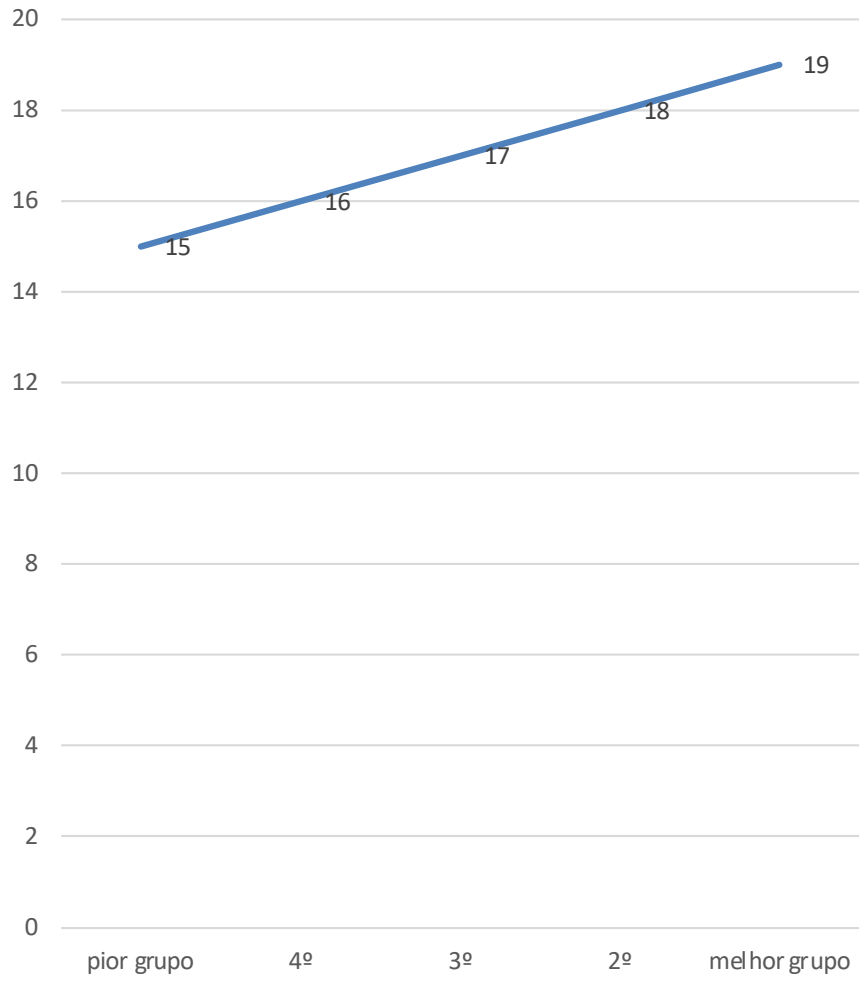
- goal
  - predict if a loan will be paid or not
- data
  - bank loans
- groups of 3
- grade
  - (60% SA + 20% CS + 20% P) * IF
    - submitted assignment
    - competition score
    - presentation
    - individual factor
  - 50% of final grade
  - minimum grade of 7/20

evaluation: competition (2/2)

CS_i = min_j(SA_j) + CompetitionRank_i*

(max_j(SA_j) - min_j(SA_j) / #groups

– ex.
  - best is max_j(SA_j)=19
  - worst is min_j(SA_j)=16



yes, there's a meme for it

team

Rita Ribeiro

- assistant professor @ FCUP
- director MDS (FCUP)
- researcher @ INESC TEC

Carlos Soares

- associate professor @ FEUP
- external advisor for IS @ Fraunhofer AICOS
- researcher @ LIACC & INESC TEC

João Mendes Moreira

- associate professor @ FEUP
- director MDSE (FEUP)
- researcher @ INESC TEC

Nuno Moniz

- invited assistant professor @ FEUP/FCUP
- researcher @ INESC TEC

support

- moodle
  - materials
  - evaluation
  - communication
- ... also on slack
  - https://join.slack.com/t/slack-gif7267/shared_invite/zt-xk1d2c3m-sN126Bu7fCo2GdHSYCOm2A
- mail
  - csoares@fe.up.pt
  - rpribeiro@fc.up.pt
  - jmoreira@fe.up.pt
  - nuno.m.moniz@inesctec.pt
- … with subject: “[ML@M.EIC]…"

UNIVERSIDADE DO PORTO

(concepts are more important than)

software

- programming
  - R
  - Python
- … or point & click
  - RapidMiner
    - https://rapidminer.com/
    - academic licence
      - installation instructions at http://docs.rapidminer.com/studio/installation/
      - insert the license key, according to the section "Entering the key in RapidMiner Studio" at http://docs.rapidminer.com/studio/installation/manual-license.html

H4sIAAAAAAAAAF2RXXOiMBiF/8oOt9bKhyB0pjOr
+IWiqCBi172IJGgsBpoEATv+90XWbWf2vctzzpycN
/kUYhwiwtDEdebCi/C5E1KawCzkFtwJLzuBghTDM
yaINhnPIE52wtOXx0eU4YSwyvlrJ3Se5cZO+P0tDy
DmlVznoAqA+wnEjwQUIohIiCpZFyvyKDJFZUVIFsc
VYxxQ3gf8bpI0STYkVRbvU2moSDGtI78M7bYmK0
Y934FrhmilVpsRcEZ1GRPQOGE/3ARQxOo66AxwX
Gshq+nPCD1n6XPK/+2bIsrx3f3yebtVKKz25hRgwt
nf9P9AbXpUsPp1NJBRqGoiahqKLDfbahs0AdRRU1
H3MuxobS1SYH0dICTh9Wrs8RY34Ulg+EAAzyiq/m
mSZ5Km6dvLkB7PUZSzmZfnsKUcBj2+iI3zNDvrl65
Ny0hbTfSQT2UwM/xlcHTh9H01W/i9kcf6uSTRaaS
rqmu8ocWbulnlm7F1WQWtj6tpTbRiXLSc93VwW
sv2AFs+Op26WxFfzXgsBp1UmuV6B17hfEjIOHBIgs
XuRobGSBwCBbwlxGN+LHVcc5b7+gfwNuu8ExXjx
aF3WcPV3KA49MhAPGxCmfhyELdLQIuGUrj9pXXs
lXbkdhqtyXEiO63RWAoSug3S4miJ+xl24ZA5V8xE7
9JbLR0ajNrnHty3etGGZwdzbWrI6zql2IjLaDRfbe3T
8Kjul8N3W12Odc9Up2VaJHaR57ZiMos4y+0xW76
+Crc/bnxkCRIDAAA=

UNIVERSIDADE DO PORTO

bibliography

- Moreira, J. M., de Carvalho, A. C. P. L. F., & Horváth, T. (2018). *A General Introduction to Data Analytics*. *A General Introduction to Data Analytics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119296294

- Data Science for Business, by Foster Provost, Tom Fawcett
  - general book, read from end to end
- Data Mining: The Textbook, by Charu C. Aggarwal.
  - textbook, to place in the shelf and consult frequently
- Data Mining: Concepts and Techniques, 3rd ed, by Jiawei Han, Micheline Kamber and Jian Pei
  - textbook, but I prefer the previous one
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition), by Trevor Hastie, Robert Tibshirani, Jerome Friedman
  - more fundamental
- Data Mining with R: Learning with case studies, by L. Torgo
  - focused in cases