CS-E4600
Algorithmic methods of data mining

Aristides Gionis
Dept of Computer Science

Slide set 1: Course logistics

schedule — autumn 2018

lectures: mon, tue 16.15 - 18.00, period I at CS building, T1

exercises: thu 14.15 - 16.00, period I at CS building, T2

exercise sessions will not take place every week

will be announced in advance

exercises are meant for clarifications on the theory

hints for homeworks, sample solutions, etc.

one can also get this information via speaking to the teaching staff of the course

programming project during period II

teaching staff

instructor: Aris Gionis, aristides.gionis@aalto.fi

office number: B345

communication: mycourses forum, or by email

teaching assistants:

Suhas Muniyappa, suhas.muniyappa@aalto.fi

Han Xiao, han.xiao@aalto.fi

Nikita Alexandrov, nikita.alexandrov@aalto.fi

communication: mycourses forum, or by email

grader(s) will be needed; announcement will follow

communication and material

all course information available via mycourses.aalto.fi:

slides

lecture notes

reading assignments

homeworks and project

announcements

in principle one can follow and pass the course without attending the lectures

A! Aalto University

assignment

3 take-home problem-set homeworks (during period I)

1 programming project (during period II)

1 in-class final (Wed 12.12.18)

each assignment counts 20% for the final grade

to pass the course one needs to get at least 20% of the max grade in each assignment

I-don't-know policy: "I don't know" answers receive 15% of grade

late-day policy: 5 late days (weekends count)

Aalto University

homeworks vs. project

homeworks

homeworks will be "pen & paper" problems

can discuss with other students, but you should write your own solutions

if you discussed the solution with other students, or got hints from the internet, you should give proper credit

project

implementing and testing existing data mining algorithms

can be done in groups of at most 2

you can find your own partner or ask us to pair you

approximate schedule

week 37: homework 1 is out

week 39: homework 1 is due

week 39: homework 2 is out

week 41: homework 2 is due

week 42: homework 3 is out

week 44: homework 3 is due

week 44: programming project is out

week 51: programming project is due

week 50: final exam

Aalto University

textbook and other reading material

Leskovec, Rajaraman, and Ullman

Mining of massive datasets

Cambridge University Press and online

http://www.mmds.org/

some lecture notes and pointers to material in the internet

syllabus

introduction to data mining

distance functions and embeddings

high-dimensional data

similarity search, locality-sensitive hashing

data streams and data sketching

data clustering (k-means, k-median, etc.)

graph partitioning and spectral graph analysis

link analysis and methods for ordering data

disclaimer

the course is not about getting hands on with datasets

instead it is about principles, foundations, and algorithmic techniques, which are recurrent in data mining

we will motivate everything with applications and examples

but there will also be some level of abstraction

A! Aalto University

prerequisites

basic concepts from algorithms and data structures:

sorting, hashing, arrays, heaps, graphs, etc.

analysis of algorithms:

O-notation, dynamic programming, NP-hardness, approximation algorithms

basic probability concepts:

conditional probability, random variables, independence, expectation, binomial, Bernoulli, and other simple distributions

linear algebra:

eigenvalues and eigenvectors

programming:

ability to process data and implement simple algorithms in some programming language

A! Aalto University

prerequisites

course assumes basic computer-science and math background

however, students with different background are welcome

if you do not know something, ask

    take advantage of the exercise sessions and office hours

    ask the instructors for background reading material

    utilize the mycources mailing list

    make study groups with your colleagues

    go to the library

    use the internet

A! Aalto University

self-assessment quiz