

Lect 1 - Part 1: Intro 1

Machine Learning For Language Technology

An Introductory Course

Marina Santini

Changelog:
14 Oct 2016, 5 Nov 2016

2016

Lect 1 - Part 1: Intro 2

Machine Learning For Language Technology

PART 1: INTRODUCTION TO THE COURSE

2016

You are encouraged to point out any inconsistencies or inaccuracies you find in the course material.

3

"I get mad at my students when I make a mistake in lecture and they don't correct me. Sometimes it seems like I could inadvertently write $2+2=5$ on the board, and nobody would speak up. Maybe they would come up after class and say, "Professor, since you said $2+2=5$, is that how you want us to answer the question on an exam?"

– Ben C---- Nov 23 '13 at 0:46

Lect 1 - Part 1: Intro 2016

About the Course

4

- A gentle introduction to Machine Learning applied to Language Technology.
- The focus of the course is on the simplest ML algorithms that are used in Language Technology.
- The course includes the theoretical description of some ML methods and the practical use of these methods (as implemented in the Weka workbench) for classification purposes.
- Teaching is based on the Flipped Classroom pedagogical approach.

Lect 1 - Part 1: Intro 2016

Lecture 1 - Part 1: Required Action

5

Listen to the following talk given by [David Black-Schaffer](#) at KTH about the benefits of the Flipped Classroom pedagogical approach.

Lect 1 - Part 1: Intro 2016

Course Website & Contact Info:

6

- http://stp.lingfil.uu.se/~santinim/ml/2016/ml4lt_2016.htm
- Contact details:
 - santinim@stp.lingfil.uu.se
 - marinasantini.ms@gmail.com
 - marinaromestockholm@gmail.com

Lect 1 - Part 1: Intro 2016

Students' Email addresses

7

- Make sure you send your email addresses to me.
- Send me the email address that you check daily.
- The communication via email is important for this course.
- Once I got your email, you will receive my reply and the enrolment key to access the Scalable Learning platform.

Lect 1 - Part 1: Intro 2016

Outline

8

1. Attendance
2. Examination
3. Course Organization
4. Course Content
5. The Flipped Classroom
6. Expected Learning Outcomes

Lect 1 - Part 1: Intro 2016

9

Attendance

Lect 1 - Part 1: Intro 2016

Attendance is Mandatory

10

- There is a mandatory 80% attendance requirement for both:
 - presentations delivered through the online platform
 - in-class lab sessions
- The whole course is made of 10 lectures.
- 80% attendance means that you can miss a couple of lectures
- In other words, you should attend 8 lectures out of 10 to be **eligible** to pass the course

Lect 1 - Part 1: Intro 2016

More than 20% absence may result in additional assignments

11

- Plan carefully, if you already have commitments.
- Notify me if you already know that you cannot comply with the 80% attendance requirement.

Lect 1 - Part 1: Intro 2016

Definition of "lecture"

12

- A lecture = out-of-the-class learning + in-class practice
- If you do not attend the out-of-the-class part, there is no point in attending the matching lab session

Lect 1 - Part 1: Intro 2016

Attendance

13

- Not valid: out-of-the-class learning=**no**; lab=**yes**;
- Not valid: out-of-the-class learning =**yes**; lab=**no**;
- Valid**: out-of-the-class learning =**yes**; lab=**yes**.

3	14/11	10:15-12:00	9-2042 (Turing)	Online: Decision Trees (slides: 1, 2) In-class: Lab	- Handout - Daume' III (2015: 10-18) - Witten et al. (2011: 99-108)
4	21/11	13:15-15:00	9-2043 (Chomsky)	Online: Evaluation (slides: 1, 2) In-class: Lab	- Handout - Daume' III (2015: 60-67) - Witten et al. (2011: Ch 5)
5	24/11	10:15-12:00	9-2042 (Turing)	Online: k-Nearest Neighbours (slides: 1, 2) In-class: Lab	- Handout - Daume' III (2015: 26-32, excl. 2.4) - Witten et al. (2011:131-138)
6	28/11	13:15-15:00	9-2043 (Chomsky)	Online: Naive Bayes (slides) Feature Selection (slides) In-class: Lab	- Handout - Daume' III (2015: 53-59; 107-110) - Witten et al. (2011: 90-99; 305-308; 314-315; 322-323; 328-329; 331-332; 334)
7	05/12	13:15-15:00	9-2043 (Chomsky)	Online: Perceptron (slides) Generative vs Discriminative (slides) In-class: Lab	- Daume' III (2015: 39-52)

Lect 1 - Part 1: Intro 2016

Examination

14

Lect 1 - Part 1: Intro 2016

Examination: 3 home assignments

15

- The course is examined by means of 3 home assignments.
- The 3 assignments have equal weight.
- Each assignment will be graded with the following marks:
 - Underkäänd (U) [Fail]
 - Godkänt (G) [Pass]
 - Val Godkänt (VG) [Distinction]
- In order to pass the course with G, a student must get a G on each home assignment.
- In order to pass the course with VG, a student must get a VG on at least two home assignments.
- If a student fails the examination, additional assignments will be required in order to receive a passing grade on the course.

Lect 1 - Part 1: Intro 2016

Assignments: Weka Workbench

16

1. Supervised Learning: Decision Trees & k-NN
2. Supervised Learning: NB (Att: the video says also "Perceptron" but it will be only NB 😊)
3. Unsupervised Learning: Clustering

Assignments might include theoretical questions.

CREATIVITY IS STRONGLY ENCOURAGED !

Show that you understood what you are doing by interpreting data and classifiers' results. Suggest improvements or novel solutions.

Lect 1 - Part 1: Intro 2016

Cheating

17

- Any assignment that is handed in must be your own work.
- However, talking to one another to understand the material better is strongly encouraged: recognizing the distinction between cooperation and cheating is very important!
- COOPERATION with other students IS WARMLY ENCOURAGED!**
- Plagiarism—copying from others—is condemned and measures will be taken if it happens.

Lect 1 - Part 1: Intro 2016

Students' Responsibilities

18

- Attendance (virtual and physical)
- Reading
- Submission of 3 home assignments.

Lect 1 - Part 1: Intro 2016

Students MUST read 😊

19

- Listen to online presentations, reply to the quizzes AND **read** the pages associated with the lectures (see the course website).

Lect 1 - Part 1: Intro 2016

20

Course Organization

Lect 1 - Part 1: Intro 2016

Math and Statistics

21

- This course builds upon the math and statistics you learned in the math course.

Lect 1 - Part 1: Intro 2016

Computers

22

- Either use your laptop (recommended) or use the computers in the classroom.
- The lab tasks are based on the Weka software package, which means that:
 - You should be able to install the software and deal with any issues about memory limits or any other software or hardware problems (troubleshooting).
 - Choose the computer that you wish, but keep in mind that YOU are responsible for the troubleshooting.

Lect 1 - Part 1: Intro 2016

Lab Tasks

23

- At each lab session, you will be given the lab tasks of the day.
- Lab tasks are based on the assumption that:
 - You have listened to and understood the online presentations before coming to the lab.
 - You have read the required reading before coming to the lab.

Lect 1 - Part 1: Intro 2016

Interaction during in-class labs: activation of the knowledge

24

- This is the ACTIVATION phase.
- I will help as LITTLE as possible during the labs, since the interaction with pairs is important for the activation of your knowledge.
- You will work in random groups on the lab tasks and present your solutions and reflections in an open discussion setting.

Lect 1 - Part 1: Intro 2016

Lab Sessions: 3 parts

25

1. Short feedback on the online presentation and quizzes. *Timeframe: approx 10 min.*
2. You will be given lab tasks to complete. You and your team will work to complete the lab tasks. *Timeframe: up to the end of the first hour.*
3. During the second hour, your group will present your solutions and discuss them with me and the audience.

Lect 1 - Part 1: Intro 2016

Lab Tasks: Example

26

Lecture 02: LAB Assignment

Weka: Data Exploration and Preprocessing

ACKNOWLEDGEMENTS: THIS LAB ASSIGNMENT IS BASED ON THE CONTENT OF THE WEKA BOOK. TASKS HAVE BEEN BORROWED AND ADAPTED FROM MARTIN D. SYKORA'S TUTORIALS (<[HTTP://HOME.PAGES.LBORO.AC.UK/~COMD52/COC131/](http://homepages.lboro.ac.uk/~comd52/coc131/)>).

Required Reading for this Lab Assignment

- Witten et al. (2011): Ch 2; Ch 10; Ch 11: 407-410; Ch 17: 559-562.

ATT: datasets can be downloaded from here:
<<http://ftp.ingff.usw/~samim/ml/2015/datasets/>>

Free material

Free Weka book (2005), 2nd edition:
<<http://home.etrs/~vni/os/dmsw/Morgan.Kaufman.Publishers.Weka.2nd.Edition.2005.elsevier.pdf>>

Learning objectives

In this lab assignment you are going to:

- install the Weka software on to your computer;
- explore data in the form of an arff dataset;
- preprocess data using statistical transformations;
- visualize data in their different formats.

Preliminaries

The benefits of interacting with peers: Peer Learning

28

- ▣ Cooperating with others to optimize your understanding of the topic
- ▣ Fostering independent-thinking
- ▣ Enhancing problem-solving skills.
- ▣ Finding the best way to show that you master the topic of the day both practically and theoretically.

Lect 1 - Part 1: Intro 2016

29

Course Content

Lect 1 - Part 1: Intro 2016

Group interaction

27

- ▣ You contribute to the success of your group. The aim is to solve as many subtasks as possible.
- ▣ Be ready to present, explain and motivate your solutions in front of the other groups.
- ▣ You and your group can interact with other groups. However, your solutions should be original (whenever possible), worked out and agreed with your OWN group.

Lect 1 - Part 1: Intro 2016

Content

30

1. Introduction
2. Basic concepts of ML
3. Decision Trees
4. Evaluation
5. k-Nearest Neighbour
6. Naive Bayes
7. Perceptron
8. k-Means
9. Hierarchical Clustering
10. Wrap-up

Lect 1 - Part 1: Intro 2016

Reading: only chapters specified in the course website

31

IMPORTANT: Read the handouts carefully. The handouts are the transcripts of the online presentations **plus** additional information and/or additional links that might be of interest.

> Handouts

> Hal Daumé III (2015). **A Course in Machine Learning**. Copyright © 2015.

Ian H. Witten, Eibe Frank, Mark A. Hall (2011). **Data Mining: Practical Machine Learning Tools and Techniques**. 3rd Edition. Morgan Kaufmann Publishers.

> Pedro Domingos (2012). **A Few Useful Things to Know about Machine Learning**. Communications of the ACM, 55(10).

Lect 1 - Part 1: Intro 2016

32

Flipped Classroom

Lect 1 - Part 1: Intro 2016

What is a "flipped classroom"?

33

□ Short answer: The flipped classroom inverts traditional teaching methods, delivering theoretical knowledge online outside of classroom and moving exercises into the classroom.

□ The traditional classroom is teacher-centered, a Flipped Classroom is student-centered.

Lect 1 - Part 1: Intro 2016

Flipping learning is upside down

34

□ The basic idea is to reverse the structure of traditional teaching.

□ Traditional teaching usually is based on:

- lectures that are delivered in a classroom by a lecturer
- homework carried out by students by themselves, not in the classroom

□ With the flipped approach, we will do the opposite:

- you will listen to the online presentations at home
- you will be in the classroom to do your homework (that we will call lab sessions)

Lect 1 - Part 1: Intro 2016

The flipped classroom:

35

Hybrid learning approach: virtual and physical attendance

- Online presentations → Scalable Learning platform
- Lab sessions → in-class

Lect 1 - Part 1: Intro 2016

Why is it beneficial?

36

□ individualized learning for students

□ students can move at their own pace and review what they need when they need

□ ability for students to catch up on missed lessons easily through the use of video clips

□ etc.

Lect 1 - Part 1: Intro 2016

Open issues

- A Flipped Classroom's success is dependent on student participation.

Lect 1 - Part 1: Intro 2016

Difficulties

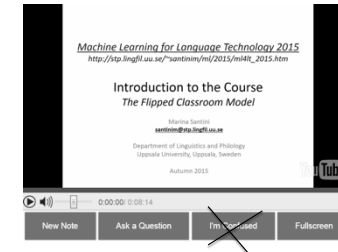
- Preparing a flipped course takes a very long time.
 - Re-Use
 - Practical parts are hard to balance with the theory and with the team work
 - etc.

.... explorations ...

Lect 1 - Part 1: Intro 2016

The Scalable Learning Platform

- We will use platform that has been developed in Sweden (by Swedish Institute of Computer Science and Uppsala University) and it is called Scalable Learning.
- Create your own account and sign up for the course using the enrolment key that will be sent to you.



Lect 1 - Part 1: Intro 2016

Online Presentations, Video Clips and Quizzes

- A presentation is made of several video clips.
- The length of the presentations and the length of video clips are variable.
- The number of the quizzes per presentation is variable. **Quizzes are NOT graded.**

Lect 1 - Part 1: Intro 2016

Scalable Learning at Uppsala Uni

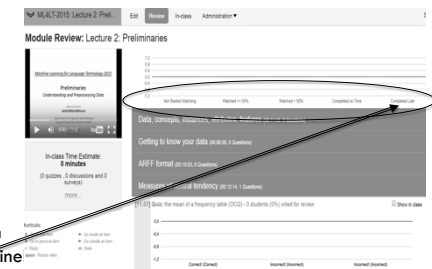
- The platform is already successfully used at Uppsala University.
- David Black-Schaffer (Department of Information Technology, UU) is regularly using it for his own courses.

Watch David's video presentation for motivation, aims, and outcomes.

Lect 1 - Part 1: Intro 2016

Scalable Learning: your duties

- Virtual attendance
- Listen to the video clips
- Answer the quizzes
- Read the associated chapters
- Each presentation has a deadline



Communication and Interaction

43

- Ask questions through the platform or by email: either you will receive an individual answer or we will discuss your questions in class.

Lect 1 - Part 1: Intro 2016

44 Expected Learning outcomes

Lect 1 - Part 1: Intro 2016

Expected Learning Outcomes

45

1. Apply basic principles of machine learning to natural language data;
2. Show theoretical and practical knowledge of:
 - Supervised models
 - Unsupervised models
 - Data representation and preprocessing
 - Evaluation of the results
3. Use of a standard machine learning package for practical classification and evaluation (the Weka workbench)

Weka logo, featuring weka, a bird endemic to New Zealand

Lect

Summary: basic stuff

46

- 80% attendance requirement
- One lecture = online presentation + matching lab session
- Lab tasks & quizzes are not graded
- Examination: In order to pass the course successfully, at least 3 Gs are needed
- Flipped Classroom & Scalable Learning
- Cooperation is encouraged, cheating is condemned.

Lect 1 - Part 1: Intro 2016

Questions?

47

Lect 1 - Part 1: Intro 2016