ML4LT - 01a: Introduction to the course

Changelog: 3 Nov 2016, 7 Nov 2016

Lecture 1 – Part 1

Welcome to the Machine Learning for Language Technology course. This course is a introduction to ML for undergraduate students.
Lecture 1 is divided into 2 parts. The first part is about the rules, the strategy and topics of the course. In the second part, we will talk about Machine Learning as a discipline, how it is defined and in which scenarios can be used.
This course is a very gentle introduction to Machine Learning applied to Language Technology. It has been designed for undergraduate students in linguistics and computational linguistics.
The focus of the course is on the simplest ML models that are used in Language Technology. It is gentle because we will not go through ALL the mathematical details and we will focus only on the **main** mathematical components of each ML method. The course requires commitment and regularity. The degree of commitment depends on your educational background and, of course, in your personal interest in the course. The course includes the theoretical descriptions of some ML methods and the practical use of classification models for classification tasks. For the practical part, we are going to use Weka workbench. Teaching is based on the Flipped Classroom pedagogical approach.
This course requires the knowledge of basic concepts of probability and statistics (see the math course). The course requires also some patience. Some of the concepts are complex and require some time to fully grabbed.
Each lecture has a required action or some required reading to be done. The required action for this part of Lecture 1 is to listen to the presentation by David Black-Schaffer in order to get a full grip of the benefits of the flipped classroom approach.

Attendance

Attendance is mandatory. If you do not meet this requirement, you will not be entitled to pass the course, even if you submit the 3 home assignments (see Examination below). 80% means that you are required to attend at least 8 out of 10 lectures.
If you have more than 20% absence score and you want to pass the course you will be required to hand in additional assignments. The number of assignments depends on the quality and the length of your absence. Plan very carefully if you know in advance that you can't attend all the lectures, and notify me.
A lecture is a combination of: out-of-the-class learning & in-class practice. It is NOT either one OR the other. It is both. If you do not attend the online presentation before the lab, there is no point in attending the lab because the lab based on notions that you have learned with the online lecture.

Examination

Attendance is one requirement to pass the course. The other requirement is to submit and pass 3 home assignments. These 3 home assignments are already online. You can start working on them as soon as you feel ready. The deadline is for all of them Sunday 18 December 2016, 23:59. You can submit earlier than that. Later submissions are not accepted. You will get a grade on the course week 51. If there are special problems, we can discuss them with Mats who is the course administrator. The grades are the usual ones: G and VG. If you fail one of the assignment, you will be asked to submit a additional assignments (komplettering).
The assignments are based on the weka workbench. Two of them are on supervised learning. One of them on unsupervised learning. Assignments might include theoretical questions that show that you have read and understood the required reading. In the assignments, creativity is much welcome. Show that you have understood the topics by giving sensible but personal interpretation of the data and the results.

Cheating is not allowed. It is not allowed to copy, but cooperation is very much encouraged.
The course is made of on online presentations, lab classes, slides, handouts and text books, examination via 3 assignments. Students' responsibility include attendance (physical and virtual), reading, submission of 3 assignments.
Reading is compulsory and all the required reading is spell out on the course website and in the slides.

Organization of the course

The course contains math and statistics. We are not going to re-explain what you have already learned in the math course. Only new notions will be explained.
In the lab classes we are going to use computers. Either use your own computer or the computers in class. The lab tasks are based on the weka workbench. You are responsible of the installation and troubleshooting. Usually weka is very well-behaved. But if something happen, I will ask you to find the solution of your problem and share it with your classmates. We could also build a forum to share problems and solutions. We will see. I have not done it, just because this course is very short, basically one month, and I do not know yet how a forum interaction would be effective in this very short timeframe.
We have 10 lectures: 8 lectures are based on the combination of online + in-class attendance. 1 lecture (this one) is face-to-face. 1 lecture (the last one) is only online without any practical lab. The lab session is based on the assumption that you have listened and understood the online presentation and that you have read and understood the required reading.
The lab session is the activation of your knowledge. I will help you as little as possible. I will go around and observe what you are doing. I might take part in group conversations. But the work must be done by you. If you do not understand the task, it means that you have not studied enough before coming to the lab. If you have doubts, ask your class mates. But, pls, do not passively apply what somebody else says, and do not wait for the "ideal" solution. Use your knowledge and your critical mind to address a problem. If you do not understand how a problem is formulated, use the context: what is the task about? what is the purpose of the task? etc. You will be divided in groups and you are supposed to provide answers to as many questions as possible. In the end, you will present your results.
The lab session will be divided into 3 parts: a very short feedback (it can be at the beginning or at the end of the lab session: it depends); you will be given lab tasks to complete and you and your team will work on it up to the end of the hour. in the second hour you will present your solutions and discuss them with the audience.
Lab tasks are not graded. Do not be afraid of failing. you will not get any.You must try out solutions in order to memorize what you are learning.
Each member of a group contributes to the success of the group.
The interaction with peers is extremely important within the Flipped Classroom strategy. You learn from the others, you develop your critical thinking, you enhance your problem-solving skills and try to give your own interpretation of a problem.

Topics of the course

Today I will start introducing what is usually meant by ML. Next lecture we will introduce the main concepts that characterize ML. In the practical lab you will install weka and start exploring a dataset. Then we will study 4 supervised methods, namely Decision Trees, k-Nearest Neighbour, Naïve Bayes and Perceptron. After having studied the first supervised method, ie the DTs, we will see how to evaluate the results. What you will learn in Lect 4 is valid for all the ML methods, not only for DTs. Lectures 8 and 9 are about unsupervised methods and will study k-means and hierarchical clustering. The last lecture is online (with no lab associated with it) and it is a list of things that you should take with you and always remember.

Required Reading

It is not easy to find a text book for a course like this. This course is unique in a way. There are hundreds of ML courses. But I could not find a ML book that is specifically tailored for language technology at undergraduate level.

Handouts

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.
The textbooks and the paper that have been chosen for this books are for ML in general. The first one, Daume', is a draft version of a manual that Hal Daume' has created out of his course notes at UMIACS (University of Maryland Institute for Advanced Computer Studies).
This book is unfinished and version that we use is the latest draft. You will study only some chapters (those that are indicated in the timetable and in the lecture slides). These chapters are relatively easy.
The manual by Witten and co-workers is the manual that accompanies with the weka workbench. Weka is a very documented kind of software. This manual is also relatively easy.
The paper by Domingos is useful to get a handy list that one must keep in mind when applying ML methods to any kind of data. Pedro Domingos is a lecturer and professor on machine learning at the University of Washington and author of abook titled "The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World". Domingos had a free course on machine learning online.

The flipped classroom approach

The flipped classroom reverses the traditional teaching.
It is a hybrid learning approach: as we said the attendance is virtual (on the e-learning platform) and physical.
The basic idea is that students are given time to listen and understand the theoretical lecture at their own pace and they apply or better "implement" what they have learned in class.

The role of the teacher is behind the scene. The teacher is kind of director, setting up and preparing the scenes, and the students are the actors who perform and are responsible for their own performance. Teacher's role: The teacher should be invisible on the scene because the students have been given all the tools to act independently. So, the students can freely express all their potential. They are not just sitting waiting for the teacher to give them the correct solutions. There is no correct solution in ML. Practical ML is mostly about interpretations. Interpretations of data, of algorithms behaviour, of feature selection. All interpretations are plausible if supported by reasonable arguments and good results.

Why is this pedagogical approach beneficial? Because you have the opportunity to learn at your pace and internalize what you learn.
However, the Flipped Classroom is not sheer magic. It is a pedagogical strategy. And the success of it depends on many factors. One fundamental factor is students' involvement and commitment. It students are not involved or interested in a subject, Flipped Classroom alone cannot do the miracle. Other approaches must be found.
For a teacher, the Flipped Classroom is very demanding. It takes very much time. The theoretical part and practical part might be hard to combine and balance, etc.
The Flipped Classroom will be applied via the Scalable Learning platform. You will have your own account, you will be notified when a lecture is online. You have several actions you can perform, ie re-play, asking a question, etc. I suggest NOT using the button " I am confused". I would prefer you send me a question saying what you are confused about. If you just press the button "I am confused", I can't be sure of what you are confused about.
The length of the presentations is variable. It depends of the topic. Some are long, some are short. Presentations have quizzes. Quizzes are compulsory, but they are not graded.

Please listen to David's presentation (the link is on the timetable) to know more about the Flippled Classroom and the Scalable Platform.
Remember: lectures have a deadline. If you do not listen the online lecture by the deadline, your attendance rate will be affected. You will be marked as absent and you should not attend the in-class lab.

Learning Outcomes

At the end of this course, you should show knowledge of supervised and unsupervised models, you should be able to pre-process the data and evaluate the results. You should feel familiar with the weka workbench.

Summary

- 80% attendance is a requirement of the course.
- One lecture means: online presentation + in-class lab;
- Labs and quizzes are not graded.
- The course is examined by 3 home assignments.
- The course is based on the Flipped Classroom and Scalable Learning.
- Cooperation is encouraged, cheating is condemned.

Tips:

Plan ahead

To complete this course, you must pass all assignments Review your upcoming assignments, especially the first assignment, to plan enough time.

Use due dates to stay on track

The course moves quickly, so it's important to stay on track. Due dates will help guide your progress,

Get help from your classmates

If you're stuck or want to discuss the course material, talk to your classmates.

Contact me for any problem

If you think that after having listened to lectures, heaving read the required reading, having talked to your classmates, you still do not have a full grip on a topic, contact me and will try to understand what went wrong in the learning process, why you are having difficulties in understanding. It can be of background knowledge, it can be … anything. Just let me know.

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

-- the end --