Draft Syllabus

1. Course description.

a Title of a Course : **Statistical learning theory** (for Bachelor level)

Lecturers: Bruno Bauwens and Queintin Paris

b Pre-requisites : Linear algebra and probability theory.

c Course Type : Elective.

d Abstract : In this course we study the theoretical framework that inspired the development of two important families of machinelearning algorithms: support vector machines and boosting. In a typical classification task we are given (1) a dataset for training and (2) a set of classifiers [for example, neural networks of some size, or polynomial threshold functions of some degree]. The learning algorithm should use the training dataset to select one of the classifiers from the set so that the selected function performs well according to some learning criteria. On one side, we want to have a large set of classifiers to model all structure in the training data. On the other hand a large set of predictors might lead to overfitting. In this course we learn how to quantify such errors and how to apply them on many machinelearning algorithms. The bounds in turn, inspire new algorithms and we pay special attention to boosting algorithms such as AdaBoost and DeepBoosting. These algorithms are currently very succesful in many application areas.

2. Learning Objectives

- Understanding basic concepts from statistical learning theory.
- Theoretical understanding of why some machine learning algorithms are succesful in a large range of applications.
- Training of mathematical skills such as abstract thinking, formal thinking and problem solving; with special emphasis on statistics.

3. Learning Outcomes

- Knowledge of several paradigms in statistical learning theory to select models (Structural risk minimization, Maximal likelihood, Minimal Description Length, etc.)

- Calculate sizes of trainingsets for several machinelearning tasks in the context of PAC-learning (and hence calculate VC-dimensions)
- In depth understanding of boosting algorithms and a few other algorithms for machinelearning .
- Understand the link between cryptography and computational limitations of statistical learning.
- Theoretical understanding of several online learning algorithms and learning with expert advice.

3. Course Plan

1. Probably approximately correct learning

Several examples of simple learning algorithms are studied such as for learning of axis aligned rectangles. The general problem is formulated. Instance classes that are not learnable are discussed.

2. VC-dimensions

We define VC-dimensions and prove the fundamental theorem of statistical learning theory: if the VC-dimension of the set of classifiers is finite, then we can learn a correct classifier with any precision using finitely many samples. Moreover, the theorem gives us upperbound for the samples in terms of the VC-dimension and the parameters of our precision measure.

3. Structural risk minimization

We consider a class of problems that is learnable in the sense above, and relax our criteria for learning to what is called *nonuniform learnability*. We provide a characterization of classifier spaces that are learnable in this sense and again this theorem gives us sample bounds. We discuss a special framework called Minimum Description Length principle and discuss some applications.

4. The time complexity of learning and cryptography

A typical cryptographic encoding of a string is an example of a an object with a clear structure, but of which no learning algorithm can find the structure of the object in a reasonable time. In this part we discuss some learning tasks that are impossible for computational reasons. Under a plausible computational complexity assumption (which is required for secure RSA encryption) one can show that neural networks of small depth and regular languages can not be learned.

5. Boosting

A *weak learning* algorithm can generate from a train set a model that is slightly better than random guessing. We observe that are weak learnable coincide with the classes that are PAC-learnable and present some efficient algorithms to transform weak learners to the stronger PAC-learning algorithms. We study AdaBoost and DeepBoost, two algorithms that are currently very popular in machine learning. Furthermore we prove performance guarantees and given an alternative explanation of their success using game theory.

6. Online learning

In online learning there is initially no training set. After each prediction, the class of the label is declared and the learning algorithm can use this information to imporve the prediction model. We study 1) prediction with expert advice, 2) linear classification algorithm such as the perceptron algorithm and 3) the connection between online learning and game theory.

4. Reading List

a) Required: Notes provided by both teachers.

b) Optional:

- Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. "Foundations of machine learning (adaptive computation and machine learning series)." (2012).
- Kearns, Michael J., and Umesh Virkumar Vazirani. *An introduction to computational learning theory*. MIT press, 1994.

Grading System

20% for homework + 40% for intermediate exam + 40% for final exam.

The exercise part of the (intermediate) exam will be open book: the students can bring the lecture notes provided by the teachers and handwritten notes.

Guidelines for Knowledge Assessment

Methods of Instruction

The materials are presented through lectures and exercises are solved during seminars.

Special Equipment and Software Support (if required)