MACHINE LEARNING
a.y. 2017-18

Instuctor:
Prof. Alessandro Chiuso
Dipartimento di Ingegneria dell'Informazione
Via Gradenigo 6/a, 35131 Padova
e-mail: alessandro.chiuso@unipd.it , web: http://automatica.dei.unipd.it/people/chiuso.html
Office hours: by appointment (e-mail or after class)
Course web page: see the "moodle" page

Aims & Syllabus

- **Objectives:** The aim of this course is to provide the fundamentals and basic principles of the learning problem as well as to introduce the most common algorithms for regression and classification. Both supervised as well as unsupervised learning will be covered, with possibly a brief outlook into more advanced and modern topics such as sparsity and boosting. The course will be complemented by hands-on experience through computer simulations.
- **Content:** Motivation; components of the learning problem and applications of Machine Learning. Supervised and unsupervised learning.

  PART I: Supervised Learning

  Introduction: Data Classes of models - Losses Probabilistic models and assumptions on the data Models, Losses and the regression function. Regression and Classification When is a model good? Model complexity, bias variance tradeoff/generalization (VC dimension generalization error)

  Least Squares, Maximum Likelihood and Posteriors.

  Models for Regression Linear Regression (scalar and multivariate) Regularization Subset Selection Linear-in-the-parameters models, Regularization. Classes of non linear models: Sigmoids, Neural Networks Kernel Methods: SVM

  Models for Classification Logistic Regression, NN, Nave Bayes Classifier, SVM

  Validation and Model Selection Generalization Error, Bias-Variance Tradeoff, Cross Validation. Model complexity determination

  PART II: Unsupervised learning Cluster analysis: K-means Clustering, Mixtures of Gaussians and the EM estimation Dimensionality reduction: Principal Component Analysis (PCA)

Prerequisites & Background material: in the first lectures we shall review some basic material from Probability & Statistics that will be needed throughout the course. If you do not feel comfortable with this material you are strongly recommended to carefully review this material; please do not hesitate to ask questions, clarifications and/or references, but keep in mind that the remaining part of the course will strongly rely on understanding this background material. We shall also need material from your Linear Algebra course (vectors and matrices, vector spaces, linear transformation, image and kernel, eigenvalues and eigenvectors). Please carefully review this as well.

Labs: The course will be complemented with 3 Labs (not compulsory, but **highly recommended**). The labs will allow to practice with hands-on experience using real data and will be based on the language Python (a brief introduction will be provided, see below). The homeworks will be evaluated assigning, in total, a grade between **0** and **3 points** (with partial points as well). This will be **ADDED** to the grade of the written test (for instance, if your grade on the written test will be 24 and your grade on the homework will be 2.7, your final grade will be 26.7, that is 27; the final grade is rounded to the closest integer).

The following schedule will be followed (note that the topics of the labs may vary).

1. Tuesday October 24th, Intro to Python (Room Te-Ue)
2. First Lab Experience *(Linear models for regression and classification)*
   - Tuesday October 31th: first homework will be released
   - Tuesday November 7th: we shall skip lecture, lab assistance for those interested (Room Te-Ue)
   - Tuesday November 14th: first homework due.
3. Second Lab Experience *(Regularization)*
   - Tuesday November 28th: second homework will be released
   - Tuesday December 5th: we shall skip lecture, lab assistance for those interested (Room Te-Ue)
   - Tuesday December 12th: second homework due.
4. Third Lab Experience *(Non-linear models and unsupervised learning)*
   - Tuesday December 19th: third homework will be released
   - Tuesday January 16th: we shall skip lecture, lab assistance for those interested (Room Te-Ue)
   - Tuesday January 23rd: third homework due.

Course Language: the official language for the course is English. Lectures and the material are in English, and the solution to homeworks and to the final exam are expected to be in English.

Main suggested book:

- Shalev-Shwartz, S. and Shai Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014.

Additional readings (suggested):

- T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer, 2008.
- C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- K.P. Murphy, *Machine Learning A Probabilistic Perspective*, MIT Press, 2012.
- Yaser S.Abu-Mostafa, M. Magdon-Ismail, H. Lin, *Learning from Data*, AMLBook, 2012.

Final test & grading:

- **Lab experience** (*not compulsory*): three homeworks (see above). Up to **3** points as a *bonus* on the written test grade.
- **Written test**: see sample tests from last year (the written test will be graded on a scale from **0** to **30L**).
- **Final Grade**: the final grade will the the **SUM** of the grades obtained in written test and the lab experiences.

Written tests:

- I session: Monday, January 29th, 2018, 3.00pm, Rooms Ke,Ve
- II session: Monday, February 12th, 2018, 3.00pm, Rooms Ke,Ve
- III session: Tuesday, June 26th, 2018, 3.00pm, Rooms, Ke,Ve
- IV session: Monday, September 17th, 2018, 3.00pm, Rooms, Ke,Ve