UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	
ACADEMIC YEAR	
ANNO ACCADEMICO EROGAZIONE	
SUBJECT	
CODE	
SCIENTIFIC SECTOR(S)	
HEAD PROFESSOR(S)	CHIODI MARCELLO — Professore Ordinario — Univ. di PALERMO
OTHER PROFESSOR(S)	CHIODI MARCELLO — Professore Ordinario — Univ. di PALERMO SOTTILE GIANLUCA — Ricercatore a tempo determinato — Univ. di PALERMO
CREDITS	
PROPAEDEUTICAL SUBJECTS	
MUTUALIZATION	
YEAR	
TERM (SEMESTER)	
ATTENDANCE	
EVALUATION	
TEACHER OFFICE HOURS	**CHIODI MARCELLO** Tuesday 15:00 17:00 stanza del docente (edificio 13); eccezionalmente su teams Friday 12:00 13:00 stanza del docente (edificio 13); eccezionalmente su teams **SOTTILE GIANLUCA** Monday 10:00 12:00 Ufficio del docente Wednesday 10:00 12:00 Ufficio del docente

DOCENTE: Prof. MARCELLO CHIODI

PREREQUISITES	Knowledge of the foundations and methods of classical statistical inference (at the level of Statistical Inference STAD) and of inference on Linear Models (at the level of Linear Models STAD); ability to use the statistical programming environment R at intermediate level, or other scientific opensource languages (e.g. Python)
LEARNING OUTCOMES	Knowledge and understanding 1. Knowledge of advanced methods of classical statistical inference (based on the likelihood approach). 2. Knowledge of basic methods of Bayesian inference. 3. Understanding of the theoretical justifications of methods and techniques learnt in previous courses. Applying knowledge and understanding 1. Ability to specify the statistical model with a critical approach, starting from the study objectives. 2. Ability to use in an integrated way the knowledge acquired in previous courses to deal with real application problems, including non-standard ones. 3. Ability to derive theoretical results in a formal way. Making judgements 1. Critical understanding of features, potentials and limitations of statistical models already known, and ability to enrich them with extensions and new features when needed. Communication 1. Ability to discuss the characteristics of a given inferential problem, both with other statisticians and with non statisticians. 2. Ability to write a scientific-technical report, focussed on the statistical model chosen to cope with a real problem and on the subject-matter interpretation of the results. Lifelong learning skills 1. Ability to use the advanced notions acquired in successive Statistics and Applied statistics courses and for the final thesis. 2. Ability to consult and understand the international statistical literature, in order to update knowledge and technical skills.
ASSESSMENT METHODS	Final written and oral exams. The written exam consists in the analysis of a real dataset at the computing lab, using the statistical programming environment R. The candidate usually has three hours available, at the end of which (s)he must hand in a technical report. The written exam has only two possible outcomes: "Admitted to the oral exam" vs. "Not admitted to the oral exam". The necessary requirement for passing this written exam is that the candidate shows a sufficient ability: (i) to use in an autonomous and critical way the statistical methods learnt in class for analysing the specific problems characterising the proposed dataset; (ii) to interpret the statistical results found; (iii) to write a technical report. The oral exam, which can only be taken by the students who have passed the written test, consists of two phases: (i) discussion of the final technical report handed in by the candidate at the end of the written test; (ii) assessment of the knowledge and ability of the candidate to illustrate and discuss the main theoretical results taught in the front classes. In case of success, the final grade (expressed in the 18/30 - 30/30 range, plus the possible "laude") will mirror: (i) the global level of achievement, in the written exam, of the "Learning outcomes", with particular reference to those listed in the previous section under the entries sub. 2 and 4.2 (up to a max of 15/30); (ii) the global level of achievement, in the oral exam, of the "Learning outcomes", with particular reference to those listed in the previous section under the entries 1.1, 1.2, 1.3, 2.3, 4.1 (up to a max of 15/30). The final grade will be obtained by summing the two components. To pass the exam, i.e. to obtain a grade not below 18/30, the student must show a SUFFICIENT level of achievement of the "Learning outcomes" in both the written and oral exam. To achieve the maximum grade of 30/30, the student must show an OPTIMAL level of achievement of the "Learning outcomes" in both the written and oral exam. The "laude" is reserved to the students who show an excellent mastering of the course contents and a remarkable level of critical approach in their use.
TEACHING METHODS	Front class teaching, computing lab tutorials, analysis of real case studies.

MODULE STATISTICAL MACHINE LEARNING *Prof. GIANLUCA SOTTILE*	
SUGGESTED BIBLIOGRAPHY	
Lecture notes made available by the professor on the University portal. Online resources indicated by the teacher during the course. - B. Boehmke, B.M. Greenwell (2020). Hands-On Machine Learning with R, First Edition, CRC Press Taylor & Francis Group (Chap. 1, 2, 4, 5, 8, 9, 10, 11, 12, 13, 14, 17, 20, 21) - A. Ghatak (2017). Machine Learning with R, Springer (Chap. 1, 3, 4, 5) - S.V. Burger (2018). Introduction to Machine Learning with R, Publisher(s): O'Reilly Media, Inc. (Chap. 2, 3, 4, 5, 6, 7, 8) - E. Alpayidin, F. Bach (2014). Introduction to Machine Learning, Fourth Edition, The MIT Press (Chap. 1, 2, 6, 7, 9, 11, 13, 18) - C. Lesmeister, S.K. Chinnamgari (2019). Advanced Machine Learning with R, Packt Publishing Limited (Chap. 2, 3, 5, 6, 7, 8, 9, 10) - A. Kassambara (2017). Machine Learning Essentials: Practical Guide in R. STHDA (Chap. 3, 5, 12, 13, 14, 21, 24, 26, 27, 28, 29, 31, 32, , 33, 34, 35)	
AMBIT	70296-Formazione matematico-statistica
INDIVIDUAL STUDY (Hrs)	108
COURSE ACTIVITY (Hrs)	42
EDUCATIONAL OBJECTIVES OF THE MODULE	
This course represents a point of contact between non-linear models and the corresponding learning algorithms to solve problems of the following types: - Classification; - Grouping; - Regressions; - Prediction. Each of these problems requires an approach that can be different in learning. Thus, the course begins by reviewing the basic concepts and definitions of the different types of machine learning: supervised, unsupervised, semi-supervised, active, and reinforcement learning. Next, we will learn how to evaluate the results of these problems, the different metrics that exist, the need to partition datasets to ensure acceptable performance and possible improvements that may arise, such as boosting or bagging techniques to generate ensembles. Eventually, students can develop the most appropriate machine learning model, combining different types and interpreting the results of the provided solution in a multidisciplinary environment.	

SYLLABUS

Hrs	Frontal teaching
2	Preliminary topics - Definitions: sample, model, variable, algorithm, dimensionality, ... - Standardization and coding - Variables selection - Variables extraction: Decomposition into singular values (SVD), Principal Component Analysis (PCA), ...
4	Approaches to learning and related problems - Supervised learning: classification problems; and regression and prediction problems - Unsupervised Learning: K-Means and hierarchical clustering - Semi-supervised approaches: semi-supervised and active learning - Reinforcement learning: optimisation problems
4	Model assessment - Overtraining and overfitting - Data set splitting: K-fold and leave-one-out - Assessing the performance of machine learning models: classification and regression problems - Model improvement: boosting, ensembles, and bagging
2	Support Vector Machines - Introduction - Optimum separation hyperplane - The kernel trick - Supported Vector Regressor (SVR)
2	Decision trees - Representation - Entropy and information gain - Pruning - Classification and Regression Trees (CART) - Random Forest (RF)
10	Current trends in Machine Learning - Gradient Boosting - K-Nearest Neighbours (KNN) - Artificial Neural Networks (ANN) - Deep Learning.

Hrs	Practice

18	Implementation of the models described in the theoretical lessons: - Preprocessing of data sets: standardisation, coding, variables selection and extraction; - Model assessment: K-fold, Leave-one-out, ...; - Supervised learning: classification and regression problems (e.g., linear, logistic, support vector machine, decision tree-based, random forest, gradient boosting ...); - Unsupervised learning: clustering problems; - Semi-supervised learning: active and reinforcement learning ; - Deep learning and artificial neural network.

MODULE
STATISTICAL MODELS

Prof. MARCELLO CHIODI

SUGGESTED BIBLIOGRAPHY

a)appunti di lezione (lecture notes);
b)Agresti, A., (2015) Foundations of Linear and Generalized Linear Models- Wiley eds.
c) Mc Cullagh, Nelder, (1989) Generalized Linear Models-Chapman and Hall eds.
d) Wood, S. (2006) , Generalized Additive Models_ An Introduction with R- Chapman and Hall
e) Pawitan, Y. (2001) In All Likelihood. Oxford Science Publications, Oxford

AMBIT	70296-Formazione matematico-statistica
INDIVIDUAL STUDY (Hrs)	108
COURSE ACTIVITY (Hrs)	42

EDUCATIONAL OBJECTIVES OF THE MODULE

This course aims at enriching the theoretical and applicative know-how of the student in the area of statistical modelling, discussing: 1) developments in the field of non linear regression-type models (GLM and extensions); 2) some critical aspects of classical parametric inference; 3) the basics of Bayesian inference. The theoretical part, taught in the front classes, will be complemented from the applications point of view in laboratory tutorials, carried out in the R environment. After successfully attending this course, proficient students should

be able:
(i) to specify an appropriate GLM or another model for the data at hand, making inference on it and interpreting the results;
(ii) to recognise situations where an extension of standard GLMs is needed, specify an appropriate model and make inference on it;
(iii) to have a critical approach to the modelling process;
(iv) to build up on the introductory notions on Bayesian inference.

SYLLABUS

Hrs	Frontal teaching
8	(a) recall on linear models, ordinary and general, linear predictors and design matrix configuration. Multivariate Normal Distribution. Asymptotic theory of multiparametric inference in regular cases.
4	General approaches to inference. A brief Introduction to Bayesian inference. Prior and posterior distributions; the role of likelihood. Bayesian point and interval estimation.
12	Generalized linear models. Model components and their different roles: linear predictor, link function, exponential family distributions. Numerical methods for parameters estimation (IWLS). Asymptotic properties. Residuals, diagnostic. Model comparison and selection. Computational issues.

Hrs	Practice
14	e
4	Advancements in classical statistical modelling: laboratory tutorials with R.