

Symbiotic AI: Improving the data quality and the algorithms

Teacher

Cristiano TAMBORRINO

Course Website (optional)

<https://classroom.google.com/>

Code: *biuuwcc*

Course description (min 150, max 300 words)

Text documents, conversation data, images, time series, and in general data coming from sensors are often not immediately usable and need a preliminary step of pre-processing that aims at performing a deep analysis of the data in order to reduce the present noise or to search for outliers that may influence the results of the algorithms in which this data is used as an input with the consequence of results affected by low accuracy or biased. This course will focus on the increase of the quality, explainability, and usability of the available data through machine learning techniques such as anomaly or outliers detection. Another important aspect of the course is the study of solutions against the curse of dimensionality. The recent developments in the techniques of matrices and tensors decomposition represent a fundamental tool that allows a reduction of the dimensionality of the data while maintaining the most important and critical information. The course will start with an exploration of the basic theory that characterizes the topics, then we will treat the state of the art methodologies with specific use cases. For each topic, we will analyze also some relevant code.

The course is part of the activities foreseen for the WP6.6 "Sustainability of SAI" of the Spoke 6 "Symbiotic AI" project FAIR - Future AI Research (PE00000013), Extended Partnership 01 - "Artificial Intelligence: Foundational Aspects", within the PNRR, Mission 4, Component 2, Investment 1.3, financed by the European Union – NextGenerationEU.

Course period

---to define---

SSD

MAT/08

Course References (optional)

- [1] Gene H. Golub and Charles F. Van Loan. "Matrix Computations." Johns Hopkins University Press, (2013).
- [2] Gilbert Strang. "Introduction to Linear Algebra, Sixth Edition." Wellesley-Cambridge Press ISBN : 978-17331466-7-8 (2023).
- [3] Kishan G. Mehrotra, Chilukuri K. Mohan, HuaMing Huang "Anomaly detection Principles and Algorithms", Springer, (2017).

Credits and Hours

2 credits, one of lectures (8 hours) and one of practice (15 hour), for a total of 23 hours. (to be determined and confirmed)

Exam Modality

Two alternatives are available to the student to pass this exam (Teacher(s) may choose other modalities):

- 1) Paper presentation. Students present the content of 1 paper suggested by the teacher. No groups are allowed.
- 2) Project. Students implement and experimentally validate an algorithm or its variation from a paper suggested by the teacher. Projects can be done in groups of 1-3 students, depending on the algorithm.

Teacher(s) CV

Attach pdf file "CV_Tamborrino_breve.pdf".

Teacher(s) Main Publications

Peer reviewed international journals

- Bohaienko Vsevolod, Diele Fasma, Marangi Carmela, Tamborrino Cristiano, Aleksandrowicz Sebastian Woźniak Edyta. A Novel Fractional-Order RothC Model, (2023). In Mathematics, Volume 11, 1677, doi: 10.3390/math11071677
- Cristiano Tamborrino, Francesca Mazzia. Classification of Hyperspectral Images with Copulas. (2022). Journal of Computational Mathematics and Data Science. Volume 6, 100070, doi: <https://doi.org/10.1016/j.jcmds.2022.100070>.
- Cristiano Tamborrino, Roberto Interdonato, and Maguelonne Teisseire. Sentinel-2 Satellite Image Time-Series Land Cover Classification with Bernstein Copula Approach (2022). Remote Sensing 14, no. 13: 3080 (scopus id: 2-s2.0-85143974852), doi: <https://doi.org/10.3390/rs14133080>.
- Annalisa Appice, Angelo Cannarile, Antonella Falini, Donato Malerba, Francesca Mazzia, Cristiano Tamborrino. Leveraging Colour-based Pseudo-labels to Supervised Saliency Detection in Hyperspectral Image Datasets. (2021). In Journal of Intelligent Information Systems, 57(3), (Scopus id: 2-s2.0-85112142036), doi: <https://doi.org/10.1007/s10844-021-00656-7>.
- Antonella Falini, Francesca Mazzia, Cristiano Tamborrino. Spline based Hermite Quasi-Interpolation for univariate time series. (2022). In Discrete and Continuous Dynamical Systems - S, 15 (12), (WOS:000767316200001), doi: 10.3934/dcdss.2022039.

Peer reviewed conference proceedings

- Falini A, C Tamborrino, G Castellano, F Mazzia, RM Mininni, A Appice, D Malerba. (2021) Novel reconstruction errors for saliency detection in hyperspectral images. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12565 LNCS, pp. 113–124 (Scopus id:2-s2.0-85101243771), doi: 10.1007/978-3-030-64583-0 12.
- Appice A, Lomuscio F, Falini A, Malerba D, Mazzia F, Tamborrino C. (2020) Saliency detection in hyperspectral images using autoencoder-based data reconstruction. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12117 LNAI, pp. 161–170. (Scopus id: 2-s2.0-85092087010), doi: 10.1007/978-3-030-5949161.
- Falini A, Castellano G, Tamborrino C, Mazzia F, Mininni RM, Appice A, Malerba D. (2020). Saliency Detection for Hyperspectral Images via Sparse-Non Negative-Matrix-Factorization and novel Distance Measures. IEEE Conference on Evolving and Adaptive Intelligent Systems, 12th IEEE International Conference on Evolving and Adaptive Intelligent Systems, EAIS 2020. (scopus id: 2-s2.0-85088117537), doi: 10.1109/EAIS48028.2020.9122749.

Peer reviewed extended abstract

- Tamborrino Cristiano, Mazzia Francesca. On the Classification of Hyperspectral Images with different Copula's Family (2021). ICNAAM - International Conference of Numerical Analysis and Applied Mathematics. In press.

CURRICULUM VITAE DELL'ATTIVITÀ SCIENTIFICA E DIDATTICA

Informazioni Personali

Data e luogo di nascita, 03-11-1981, Brindisi

Cittadinanza, italiana

Stato civile, libero

Indirizzo lavorativo, Dipartimento di Informatica, Università degli studi di Bari "Aldo MORO", , 70125, Bari, Italia

e-mail, cristiano.tamborrino@uniba.it

Formazione

06/2022 **PhD in Mathematics and Computer Science, curriculum Mathematics, University of Bari "Aldo MORO"**, Bari, Italy.

- PON RI 2018-2020 Ministry of Education and University of Italy.
- CUP: H94F18000260006
- Research area: Data Science.
- SSD: MAT/08
- PhD Project: "Change Detection in Remote Sensing"
- PhD Thesis: "Analysis of Remotely Sensed data with Copulas and Machine Learning"
- Supervisor: Prof. Francesca Mazzia, Prof. Annalisa Appice, University of Bari, Italy
- Final mark: Doctor Europaeus

09/2017 **Master in Data Science, Università degli Studi di Perugia**, Perugia, Italy.

- Data Mining and Machine Learning: Classification techniques (Decision Trees and its derivatives such as Random Forest and Gradient Boosting), SVM, PCA, Regression techniques, Neural Network and Deep Learning
- Data Science Tools :
 - Hadoop's architecture, core components (MapReduce, HDFS) and some tools from Hadoop Ecosystem (Sqoop, Flume, Hive and Impala);
 - Apache Spark, working with RDDs and Pair RDDs, Parallel Processing in Spark and its Persistence, Spark SQL.
- Cloud Computing : PaaS , IBM Bluemix.
- Big data Processing and Tools : Stream Analytics Hackathon using Spark and Storm with ML libraries.
- Fundamentals of Computer Science : Python programming. Business Intelligence. Visual Analytics. Data-driven Marketing.
- Visual Analytics : Treemaps, Layered and Radial Visualizations, Spring-Embedder and Sugiyama Algorithm, Data visualizations with d3js.

Master's Degree in Mathematics, Università degli Studi di Bari "Aldo MORO", Bari, Italy.

- Thesis project: "Meixner processes in the pricing of options on stock indices" (Finance)
- Supervisor: Prof. Nicola Cufaro Petroni, University of Bari, Italy
- Final Score 110/110 cum laude.
- Main modules: Probability Theory, Numerical Analysis, Econometrics, Analytical Methods in Finance, Stochastic Processes in Finance, Advanced Microeconomics and Macroeconomics, Mathematical optimization and Methods.

Specialization in programming languages, *Tms High Performance Training*, Milano, Italy.

- C, C++, java, plsql, sql, html, oracle 10

Bachelor's Degree in Mathematics, *Università degli Studi di Bari "Aldo MORO"*, Bari, Italy.

- Thesis project: "Non-linear systems of ordinary differential equations, an application in the fishing industry" (Mathematical Analysis)
- Supervisor: Prof. Giuliana Palmieri, University of Bari, Italy
- Final Score 106/110
- Main modules: Calculus, Mathematical Analysis, Geometry, Algebra; Probability Theory; Numerical Analysis; Physics; .
- Programming languages: Matlab.

Keywords - Scientific interests

- Numerical Analysis, Statistics and Probability, Scientific computing.
- Hamiltonian Monte Carlo method, Numerical integrators, Matrix factorization.
- Machine learning models, supervised and unsupervised classification.
- Dependence modeling with multivariate distributions and Copulas.

Esperienza lavorativa

- 20/02/2023– **Ricercatore a Tempo determinato RTDA**, *settore scientifico disciplinare MAT08*, Dipartimento di Informatica, Università degli studi di Bari "ALDO MORO", Bari
- 31/12/2021– **Assegnista di Ricerca**, *settore scientifico disciplinare MAT08*, Centro Nazionale Ricerche
- 19/02/2023 istituto di Nanotecnologie, CNR Nanotec, Lecce.
- 6/12/2018– **Dottorando di Ricerca in Informatica e Matematica**, *settore scientifico disciplinare*
- 6/12/2021 *MAT08*, Dipartimento di Matematica, Università degli studi di Bari Aldo Moro
- 01/09/2017– **Data Scientist Junior**, *Comune di Bari*, Studio degli Open Data a supporto della pubblica
- 01/12/2018 amministrazione
- 01/09/2015– **Docente di Matematica e Fisica**, Liceo Scientifico Ludovico Pepe, Ostuni (BR)
- 01/04/2010– **Consulente Informatico**, *Accenture*, Milano
- 31/07/2012

Publicazioni Scientifiche

Publicazioni scientifiche disponibili ai seguenti links:

- <https://orcid.org/0000-0002-5478-1792>

Interessi di ricerca

La sua attività di ricerca riguarda metodi di machine learning e data science. In particolare è coinvolto nella produzione di algoritmi per l'analisi di dati di grandi dimensioni con applicazioni a dati ambientali e a dati telerilevati sia da sensori ottici che radar. Si occupa di metodi numerici e statistici per la classificazione la clusterizzazione, la salinicy detection e la change detection. In particolare ha studiato e studia applicazioni con l'utilizzo delle Copule. Ha esperienze nel campo dei metodi Markov chain Monte Carlo per il campionamento da distribuzioni statistiche in particolare con tecniche di integrazione geometrica che preservano l'energia nel metodo

Principali Corsi e Scuole Estive

- 2019– **Metodi Matematici per la Data Science**, *Scuola Estiva*, Università degli Studi di Bari
ALDO MORO
- 2019– **Computational Mathematics, Statistics and Machine Learning**, *Università degli Studi di Pavia*
- 2013– **Sistemi Satellitari per telerilevamento e Geolocalizzazione**, *Politecnico di Bari*
- 2009– **Laboratorio su dati telerilevati Attivi e Passivi**, *CNR Bari*

Conferenze

- 2022 **COMPSTAT 2022 - 24th International Conference on Computational Statistics**, *Bologna - Italia*, Families of copulas and application in remote sensing problems
- 2021 **ICNAAM 2021 - 19th International Conference of Numerical Analysis and Applied Mathematics**, *Rhodes, Grecia*, Classificazione di immagini iperspettrali con differenti famiglie di Copula
- 2021 **INDAM DAY 2021 - Conference on Mathematics and industry**, *BARI*, Statistical copulas approach for dependence in remote sensing problems
- 2020 **IEEE EAIS 2020 - Conference on Evolving and Adaptive Intelligent Systems**, *BARI*, Saliency Detection for Hyperspectral Images via Sparse-Non Negative-Matrix-Factorization and novel Distance Measures
- 2020 **ISMIS 2020 International Symposium on Methodologies for Intelligent Systems**, *Graz*, Saliency Detection in Hyperspectral Images Using Autoencoder-Based Data Reconstruction

Partecipazione a programmi di ricerca

- 2021 **Componente gruppo di ricerca**, *progetto di ricerca industriale “CLOSE to the Earth” PON Ricerca e Innovazione 2014-2020 (codice identificativo: ARS01_00141)*. Specializzazione: *Aerospazio.*, Soggetto Capofila: DTA - Distretto Tecnologico Aerospaziale