



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO
DIPARTIMENTO DI
MATEMATICA

[iNSAM]
Unità di Ricerca
Università degli Studi di Bari Aldo Moro

Alumni Mathematica

SCIENCE – MATHEMATICAL –
-Summer School-
DATA – STATISTICS

Summer School Mathematical Methods in Data Science

ON-LINE EDITION

Dipartimento di Matematica Università
degli Studi di Bari Aldo Moro

12-16 Luglio 2021

Indice

| | |
|----------------------------|-----------|
| Introduzione | 3 |
| Programma | 4 |
| Lectio Magistralis | 7 |
| Seminari tematici | 9 |
| Sessioni Didattiche | 11 |
| Seminari Aziendali | 21 |

Benvenuti alla terza edizione della Summer School MMDS 2021

La Summer School “Mathematical Methods in Data Science” è organizzata da:

- il Dipartimento di Matematica dell’Università degli Studi di Bari Aldo Moro,
- l’Associazione di Promozione Sociale Alumni Mathematica,
- l’Unità di Ricerca INdAM (Istituto Nazionale di Alta Matematica) dell’Università degli Studi di Bari Aldo Moro.

L’evento propone un percorso formativo di cinque giorni incentrato sulle applicazioni della Matematica al Data Science (Scienza dei Dati).

Le giornate didattiche sono strutturate in modo da accompagnare i discenti nel graduale apprendimento di alcuni degli strumenti matematici più utilizzati nel contesto del Data Science. Gli argomenti affrontati durante la Scuola spaziano dall’analisi statistica dei dati, alla regressione lineare dei dati, ai meccanismi di riduzione di dimensionalità e estrazione di conoscenza latente da dati matriciali, fino all’introduzione dei concetti base del Deep learning. Durante ogni sessione didattica sono previste esercitazioni pratiche con i linguaggi di programmazione *R* e *Python*. Questi ambienti di lavoro open source sono introdotti durante la sessione preliminare di laboratorio del giorno lunedì 12 luglio 2021.

Nella giornata di apertura della scuola sono previste specifiche relazioni tematiche tenute da esperti del settore e due *Lectiones Magistrales* del Prof. Contucci, professore ordinario di fisica-matematica presso il Dipartimento di Matematica dell’Università degli Studi Alma Mater di Bologna e della Prof.ssa Morini professore ordinario di Analisi Numerica presso il Dipartimento di Ingegneria Industriale dell’Università degli Studi di Firenze. Nell’arco della settimana sono proposte ai partecipanti ulteriori relazioni tematiche tenute dalle aziende sponsor della scuola.

Nonostante la modalità online, a cui siamo stati obbligati a causa della pandemia da Covid-19, non mancheranno, inoltre, occasioni per lo scambio di idee e di opinioni.

Desideriamo, infine, ringraziare lo sponsor di questa edizione, il cui supporto è stato essenziale per l’organizzazione della scuola:



Buona scuola a tutti e non dimenticate di raccontare sui social la vostra esperienza con foto e post utilizzando l’hashtag **#MMDS21** e **#DataScienceBari**

ENJOY ☺

Il Comitato organizzatore: Nicoletta Del Buono, Flavia Esposito, Stefano Franco e Francesca Mazzia.

Programma

- Lunedì 12 Luglio 2021

| | |
|--|---|
| 08:45 - 09:30 | Apertura lavori e Saluti istituzionali |
| (giornata online con accesso aperto a tutta la comunità accademica): Prof.ssa Nicoletta Del Buono | |
| Intervengono: | |
| Prof.ssa Anna Maria Candela , Prorettore Università degli Studi di Bari Aldo Moro | |
| Prof.ssa Addolorata Salvatore : Direttore del Dipartimento di Matematica | |
| Prof. Luciano Lopez : Presidente del corso di Laurea in Matematica | |
| Dott.ssa Paola Romano : Assessore Comune di Bari, Deleghe: Politiche giovanili, Pubblica Istruzione, Università, Ricerca e Fondi comunitari | |
| Prof.ssa Valeria Ruggiero : direttore del Gruppo Nazionale di Calcolo Scientifico (GNCS) dell'I.N.d.A.M. | |
| Prof. Claudio Canuto , membro del Consiglio Scientifico dell'I.N.d.A.M – Istituto Nazionale di Alta Matematica | |
| Dott. Alfredo Abrescia : Senior Manager, Fincons Group | |
| Dott.ssa Flavia Esposito : Associazione Alumni Matematica | |
| <hr/> | |
| dalle 9:35 c.a. | Interventi tematici |
| 9:35-9:55 "The Dark Side of Data Science", Prof. Giuseppe Pirlo , Università degli Studi di Bari e consorzio CINI | |
| 9:55-10:15: "Big Data e Intelligenza Artificiale tra democrazia e dominio", Prof. Roberto Bellotti , Università degli Studi di Bari Aldo Moro | |
| <hr/> | |
| 10:15 - 10:30 | Pausa Lavori |
| <hr/> | |
| dalle 10:30 c.a. | Seconda Parte |
| Moderatore Prof.ssa Nicoletta Del Buono | |
| 10:30-11:15: "Dati, elaborazione e apprendimento automatico", Prof. Pierluigi Contucci , Università degli Studi di Bologna | |
| 11:20-12:05 "Ottimizzazione Numerica e Machine Learning": Prof.ssa Benedetta Morini , Università degli Studi di Firenze | |
| 12:05-12:15: Chiusura della giornata e saluti finali | |

| | |
|--------------------|---|
| 12:30-13:30 | Lunch Time |
| 13:30-15:00 | Sessione tecnica Flavia Esposito <i>Analisi Esplorativa dei dati con R</i> |
| 15:00-15:30 | Break |
| 15:30-17:00 | Sessione tecnica Francesca Mazzia, Antonella Falini <i>Introduzione al Python e misure di similarità</i> |
| <hr/> <hr/> | |

- Martedì 13 Luglio 2021

| | |
|-------------|---|
| 09:00-12:30 | Sessione 1 Nicoletta Del Buono <i>Analisi delle Componenti Principali</i> |
| 12:30-14:00 | Lunch Time |
| 14:00-16:30 | Sessione 1: Laboratorio Nicoletta Del Buono, Flavia Esposito <i>Analisi delle Componenti Principali</i> |
| 16:30-17:45 | Sessione 2 Francesca Mazzia, Antonella Falini <i>Tecniche per la riduzione della dimensionalità e misure di similarità in Python</i> |

- Mercoledì 14 Luglio 2021

| | |
|-------------|---|
| 09:00-12:30 | Sessione 3 Claudia Angelini <i>Introduzione al Clustering</i> |
| 12:30-14:00 | Lunch Time |
| 14:00-16:30 | Sessione 3: Laboratorio Claudia Angelini <i>Introduzione al Clustering</i> |
| 16:30-17:45 | Sessione 2 Francesca Mazzia, Antonella Falini <i>Tecniche per la riduzione della dimensionalità e misure di similarità in Python</i> |

- Giovedì 15 Luglio 2021

| | |
|-------------|---|
| 09:00-12:30 | Sessione 4 Antonio Di Crescenzo, Barbara Martinucci <i>Aspetti di base relativi alla Regressione: teoria e applicazioni</i> |
| 12:30-14:00 | Lunch Time |
| 14:00-15:30 | Sessione 4: Laboratorio Antonio Di Crescenzo, Barbara Martinucci <i>Aspetti di base relativi alla Regressione: teoria e applicazioni</i> |
| 15:30-16:00 | Presentazione Aziendale |
| 16:00-16:30 | Seminario Aziendale |
| 16:30-17:00 | Incontro Studenti Aziende |
| 17:00-17:30 | Tavola Rotonda “Data Science e Mondo del Lavoro” |

- Venerdì 16 Luglio 2021

| | |
|-------------|---|
| 09:00-13:00 | Sessione 5 Monica Bianchini <i>Introduzione alle reti neurali artificiali e reti convoluzionali per l'elaborazione delle immagini mediche</i> |
| 13:00-14:30 | Lunch Time |
| 14:30-17:30 | Sessione 5: Laboratorio Simone Bonechi <i>Introduzione alle reti neurali artificiali e reti convoluzionali per l'elaborazione delle immagini mediche</i> |
| 17:30 | Chiusura dei lavori e consegna degli attestati |

Lectiones Magistrales

Dati, elaborazione e apprendimento automatico

Pierluigi Contucci

Dipartimento di Matematica, Università degli Studi di Bologna,
pierluigi.contucci@unibo.it

Le enormi capacità di raccolta di dati della tecnologia moderna hanno favorito lo sviluppo, insieme all'accresciuta capacità di elaborazione, della tecnica di apprendimento automatico. Questa, tanto quanto i prototipi del motore agli albori della rivoluzione industriale, è ancora basata su metodi euristici che vanno approfonditi e ottimizzati in uno schema di teoria scientifica. La conferenza discuterà brevemente il ruolo delle scienze dure in questo processo e la grande sfida che attende il mondo accademico

CV breve: Pierluigi Contucci è full professor di fisica-matematica presso il Dipartimento di Matematica dell'Università degli Studi Alma Mater di Bologna. Possiede un dottorato in Matematica (ISAS, Trieste 1994), un MS in Fisica Matematica (ISAS, Trieste 1991), e una Laurea in Fisica (Università La Sapienza, Roma 1989). E' stato Postdoc presso la Princeton University (1996-1998) e Assistant Professor presso l'University of California Davis (1998-2000). E' stato Visiting Professor presso il Tokyo Institute of Technology (2007) e la New York University in Shanghai (2014) e Abu Dhabi (2017). I suoi interessi di ricerca riguardano la Fisica Matematica, la Meccanica Statistica e le applicazioni (sistemi complessi, machine learning, data science per gli studi sociali e economici). Pierluigi Contucci è autore di più di cento pubblicazioni scientifiche su giornali internazionali come Communications in Mathematical Physics, Nature Scientific Reports, Journal of Statistical Physics, Annales Henri Poincare, Physical Review Letters, due monografie Perspectives on Spin Glasses (Cambridge University Press, 2012) e Probabilità Elementare, Teoria ed Esperimenti (Zanichelli, 2008). Ha curato anche dibattiti scientifici di interesse pubblico in quotidiani come Il Sole 24 Ore e giornali come Il Mulino e Vanity Fair. Pierluigi Contucci ha organizzato diverse conferenze internazionali, tra cui un trimestre sui sistemi complessi presso l'Institut Henri Poincare (Paris 2015, Director Cedric Villani). E' stato premiato con lo Schloessman Award dal the Max Planck Society (2000). E' inoltre Direttore dell'Unità di Bologna dell'Istituto Nazionale di Alta Matematica (INdAM), membro del comitato scientifico dell'Istituto Cattaneo e membro dell'Accademia delle Scienze di Bologna. E' stato recentemente contattato (novembre 2018) dal Ministero Italiano per l'Università e la Ricerca (MIUR) al fine di aiutare nell'organizzazione delle Priorità Nazionali in Intelligenza Artificiale sugli argomenti come Higher Education e per coordinare il gruppo di Intelligenza Artificiale, Cybersecurity e Robotica per il Piano Nazionale di Ricerca (2020-2027).

Ottimizzazione numerica e machine learning

Benedetta Morini

Dipartimento di Ingegneria Industriale, Università degli Studi di Firenze,

`benedetta.morini@unifi.it`

In questa comunicazione discutiamo aspetti teorici e numerici relativi alla risoluzione di problemi di ottimizzazione che nascono nell'ambito del machine learning. Dopo aver mostrato l'origine di alcuni problemi di ottimizzazione non lineari, ed eventualmente non convessi, nel machine learning, presentiamo metodi comunemente usati in letteratura e discutiamo le tematiche più attuali di ricerca, dettate dal crescente utilizzo dell'apprendimento automatico e dalla necessità di migliorare le prestazioni dei procedimenti esistenti.

CV breve: Benedetta Morini si è laureata in Matematica nel 1992 presso l'Università degli Studi di Firenze, ha conseguito il titolo di Dottore di Ricerca in Matematica Computazionale ed Informatica Matematica presso l'Università degli Studi di Padova. Attualmente è professore ordinario di Analisi Numerica presso il Dipartimento di Ingegneria Industriale dell'Università degli Studi di Firenze.

Svolge attività di ricerca nell'ambito della risoluzione numerica di problemi di ottimizzazione con o senza vincoli, sistemi di equazioni e disequazioni non lineari, minimizzazione di somme finite con applicazioni al machine-learning, sistemi lineari e successioni di sistemi lineari di grandi dimensioni in problemi di ottimizzazione. È stata responsabile di progetti di ricerca nazionali (INDAM/GNCS) e partecipante a progetti di ricerca nazionale (PRIN, INDAM/GNCS, POR FESR).

È membro del Collegio dei Docenti del Dottorato di Ricerca in Informatica e Matematica del Calcolo, Università dell'Insubria.

È membro del Collegio Editoriale della rivista Numerical Algorithms, Springer.

È membro della Commissione Scientifica dell'Unione Matematica Italiana U.M.I., Triennio 2018–2021 e Triennio 2021-202

Seminari Tematici

- **Roberto Bellotti:** Big Data e Intelligenza Artificiale tra democrazia e dominio (pag. 9)
- **Giuseppe Pirlo:** The Dark Side of Data Science (pag. 10)

Big Data e Intelligenza Artificiale tra democrazia e dominio.

Roberto Bellotti

Dipartimento Interateneo di Fisica M.Merlin, Università degli Studi di Bari Aldo Moro
Istituto Nazionale di Fisica Nucleare (Bari)
roberto.bellotti@ba.infn.it

L'analisi di Big Data attraverso tecniche di intelligenza artificiale é oggi giorno estesa a tutti gli ambiti della vita quotidiana oltre che nella ricerca scientifica. Verranno presentati alcuni casi di studio significativi di applicazione di queste tecniche nell'ambito della medicina, dell'economia e della politica, mettendo il luce vantaggi e pericoli.

CV breve: Roberto Bellotti é professore ordinario di fisica applicata e direttore del dipartimento di fisica dell'Università di Bari. Svolge attività di ricerca nell'ambito del machine learning dai primi anni novanta. E' autore di numerose pubblicazioni scientifiche sulle più importanti riviste internazionali, tra cui Nature e Science. Collabora, come esperto in questi temi, con numerose organizzazioni, tra cui Invitalia e le Nazioni Unite.

The Dark Side of Data Science

Giuseppe Pirlo

Consorzio Interuniversitario Nazionale per l'Informatica
Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro
giuseppe.pirlo@uniba.it

Il Piano Nazionale delle Ricerche (PNR) 2021-2027 è permeato di Data Scienze. Pressoché tutti gli ambiti di ricerca si stanno aprendo a questo dominio che è diventato strategico per lo sviluppo futuro di interi settori sociali ed economici che vanno da quello sanitario a quello legato alla mobilità e alle infrastrutture, da quello della formazione a quello legato al turismo e alla cultura, solo per fare alcuni esempi. Purtroppo il tipo di approccio verso questa disciplina è ancora molto confuso e al Data Science, così come ad altri ambiti innovativi quali quelli dell'AI, Machine Learning, Intelligent Systems, etc. , viene spesso delegata la responsabilità di proporre e sviluppare soluzioni efficaci ed efficienti in un contesto spesso contraddittorio e che ha già mostrato ampi limiti. Si pensi alle criticità del modello basato sul Cloud Computing o alle problematiche legate alla sostenibilità energetica dei sistemi digitali. In questo scenario è diventato indispensabile avviare una importante riflessione anche sui limiti degli strumenti a nostra disposizione e come superarli rapidamente sia da un punto di vista teorico che applicativo, per poter rispondere al meglio alle reali esigenze di sviluppo.

CV breve: Giuseppe Pirlo, laureato nel 1986 in Scienze dell'Informazione, è attualmente Professore Ordinario di Sistemi di Elaborazione delle Informazioni. IEEE Senior Member.

Già Pro Rettore Vicario dell'Università degli Studi di Bari Aldo Moro è il Delegato alla Terza Missione e alla Sostenibilità, Membro della Giunta del Consorzio Interuniversitario Nazionale per l'Informatica (CINI), Membro del Consiglio Direttivo della Rete delle Università per lo Sviluppo Sostenibile (RUS), Membro del Consiglio Direttivo del Centro nazionale di studi per le politiche urbane (Urban@it), Membro del Consiglio Direttivo dell'Associazione Italiana per l'Informatica ed il Calcolo Automatico (AICA), e presidente AICA - sezione Puglia, Presidente del Centro di Eccellenza per la Sostenibilità dell'Università degli Studi di Bari Aldo Moro.

Giuseppe Pirlo è il Coordinatore nazionale dell'ambito "Transizione Digitale e Industria 4.0" del Piano Nazionale Delle Ricerche 2021-2027.

L'attività di ricerca di Giuseppe Pirlo riguarda diversi campi tra i quali Intelligenza Artificiale, Pattern Recognition, Sistemi Intelligenti, Document Processing, Sistemi Biometrici e architetture di calcolo ad elevate prestazioni. Su questi temi è autore/coautore di oltre 350 articoli su riviste internazionali e/o presentati a conferenze internazionali. Curatore di numerosi libri e guest editor di diversi special issue è associate editor e revisore di prestigiose riviste scientifiche internazionali.

Si occupa da molti anni di problematiche legate all'innovazione, alla sostenibilità, al trasferimento tecnologico ed alle competenze digitali.

Sessioni Didattiche

- **Sessione tecnica:** Analisi Esplorativa dei dati con R, Flavia Esposito (pag. 12)
- **Sessione tecnica:** Introduzione al Python e misure di similarità, Francesca Mazzia e Antonella Falini (pag. 13)
- **Sessione 1:** Analisi delle Componenti Principali, Nicoletta Del Buono e Flavia Esposito (pag. 14)
- **Sessione 2:** Tecniche per la riduzione della dimensionalità e misure di similarità in Python, Francesca Mazzia e Antonella Falini (pag. 16)
- **Sessione 3:** Introduzione al Clustering, Claudia Angelini (pag. 17)
- **Sessione 4:** Aspetti di base relativi alla Regressione: teoria e applicazioni, Antonio Di Crescenzo e Barbara Martinucci (pag. 18)
- **Sessione 5:** Introduzione alle reti neurali artificiali e reti convoluzionali per l'elaborazione delle immagini mediche, Monica Bianchini e Simone Bonechi (pag. 19)

Sessione tecnica: *Analisi Esplorativa dei dati con R*

Flavia Esposito

Dipartimento di Matematica,
Università degli Studi di Bari Aldo Moro
flavia.esposito@uniba.it

In questa sessione verranno introdotti i concetti preliminari del linguaggio di programmazione R. Si farà uso del software Anaconda e dei programmi preinstallati in questo per la programmazione in R, tramite esercitazioni dal vivo con i propri calcolatori, verranno illustrate le modalità di trattamento dei vari tipi di dati, di liste, vettori, matrici e dataframes. Verranno forniti gli strumenti per ispezionare un dataframe facendo uso dei dataset precaricati negli ambienti di lavoro. Infine verranno illustrati alcuni esempi di importazione e esportazione di dataset e risultati.

CV breve: Flavia Esposito è ricercatrice a tempo determinato presso il Dipartimento di Matematica dell'Università degli Studi di Bari Aldo Moro. Ha conseguito il titolo di dottore di ricerca in Matematica nel 2019 presso l'Università degli Studi di Bari Aldo Moro discutendo una tesi dal titolo "Nonnegative Matrix Factorization for Knowledge Extraction from Biomedical and other real world data". Nello specifico lavora nel campo dell'analisi numerica e dei metodi di ottimizzazione di tipo low-rank per la fattorizzazione di grandi matrici di dati. E' stata assegnista presso il Dipartimento di Ingegneria Elettrica e Informatica e titolare di borsa di studio presso l'Istituto Tumori Giovanni Paolo II, dove ha potuto continuare la sua ricerca in ambito matematico con applicazioni a contesti reali come l'ingegneria e lo studio di tumori. Ha effettuato un periodo di ricerca in Belgio presso l'Université de Mons sotto la supervisione del Prof. Nicolas Gillis, uno dei massimi esperti mondiali nell'utilizzo delle tecniche di fattorizzazioni nonnegative. E' impegnata nel campo della divulgazione delle scienze matematiche con l'associazione di promozione sociale Alumni Mathematica, della quale è presidente, ed è fermamente convinta che la matematica sia alla base di gran parte dei processi applicativi moderni e che la collaborazione e lo scambio di idee con esperti di altri settori possa aiutare a mettere questa disciplina al servizio di tutti.

Sessione tecnica: *Introduzione al Python e misure di similarit *

Francesca Mazzia e Antonella Falini

Dipartimento di Informatica,

Universit  degli Studi di Bari Aldo Moro

`francesca.mazzia@uniba.it`, `antonella.falini@uniba.it`

In una prima lezione verranno forniti i fondamentali del linguaggio Python in particolare in riferimento alle strutture e agli oggetti necessari per condurre l'analisi dei dati che ritroviamo in numerosi ambiti applicativi, dalle immagini satellitari o iperspettrali, all'analisi dei documenti testuali.

CV breve: Antonella Falini dall'agosto 2019   ricercatrice a tempo determinato presso il dipartimento di Informatica, dell'universit  "Aldo Moro" di Bari. Ha conseguito il PhD presso l'Istituto di Geometria Applicata dell'universit  Johannes Kepler di Linz (Austria), tramite una borsa Marie-Curie. Ha condotto ulteriori periodi di ricerca presso la Scuola di Computer Science di Cardiff (U.K.) e la Simplewere Ltd (U.K.). Ha effettuato un periodo di post-doc di tre anni presso l'Universit  di Firenze, tramite borse di ricerca INdAM. Il suo ambito di ricerca principale riguarda la modellistica e l'approssimazione numerica. Recentemente si occupa dello studio di nuovi modelli di clusterizzatori e di fattorizzazioni per il Big Data Analytics.

CV breve: Francesca Mazzia   professoressa ordinaria di Analisi Numerica presso il Dipartimento di informatica dell'Universit  degli studi di Bari Aldo Moro da gennaio 2018. Si   laureata in Scienze dell'Informazione nel 1989, dal 1990 ha svolto il ruolo di ricercatore e poi di professore associato presso la stessa Universit . Nel 1997 ha svolto attivit  di ricerca presso il Cerfacs (Centre Europ en de Reserche et de Formation Avanc e en Calcul Scientifique) Toulouse, Francia, vincitrice di una borsa post-doc Marie-Curie. E' autrice numerose pubblicazioni su riviste internazionali. La sua attivit  di ricerca riguarda principalmente i metodi numerici per la soluzione di equazioni differenziali ordinarie, sia ai valori iniziali che al contorno, attualmente si sta occupando di nuovi modelli di clusterizzatori e di fattorizzazioni approssimate per il BigData Analytics.

Sessione 1: *Analisi delle Componenti Principali*

Nicoletta Del Buono e Flavia Esposito

Dipartimento di Matematica,

Università degli Studi di Bari Aldo Moro

nicoletta.delbuono@uniba.it, flavia.esposito@uniba.it

L'Analisi delle Componenti Principali (Principal Component Analysis - PCA) è un metodo matematico utilizzato per estrarre informazioni da un dataset rappresentandole tramite un numero di variabili ridotto rispetto a quello di partenza. Questo metodo trasforma le variabili correlate presenti nel dataset in studio in un nuovo insieme di variabili artificiali chiamate "componenti principali (PCs)" che rappresentano la maggior parte della varianza nei dati. Nella lezione si introdurranno i concetti generali legati alla PCA, le nozioni di matrice di covarianza e correlazione dei dati, di assi principali e di rotazione dello spazio degli osservabili e i meccanismi matematici per calcolare le PCs. Si discuteranno alcune euristiche per il calcolo del numero di PCs da conservare e i meccanismi empirici per interpretare le PCs. La fattibilità della PCA sarà illustrata mediante lo studio di alcuni dataset di benchmark elaborati mediante i linguaggi open source Python e R.

CV breve: Nicoletta Del Buono è professore associato di Analisi Numerica presso il Dipartimento di Matematica, dell'Università degli Studi di Bari Aldo Moro. L'attività di ricerca è stata principalmente rivolta allo studio di metodi numerici per equazioni differenziali ordinarie, calcolo di funzioni di matrici, fattorizzazioni di matrici e loro applicazioni. Più recentemente si è occupata di metodi di approssimazione low rank per matrici di dati, concentrandosi in particolare sulle fattorizzazioni di matrici con vincoli di non-negatività e su problematiche applicative quali a esempio, l'analisi di microarray e di dati biomedici, il clustering di collezioni di documenti testuali. E' autore di diversi lavori scientifici pubblicati su riviste internazionali e ha partecipato a numerosi convegni nazionali e internazionali nonché a diversi progetti di ricerca finanziati dall'Università degli Studi di Bari, dal gruppo INDAM-GNCS, dal MIUR e dalla Fondazione Cassa di Risparmio di Puglia. Completano l'attività scientifica alcuni periodi di ricerca all'estero svolti presso la Bath University (U.K.) e la North Carolina State University, Raleigh (USA).

CV breve: Flavia Esposito è ricercatrice a tempo determinato presso il Dipartimento di Matematica dell'Università degli Studi di Bari Aldo Moro. Ha conseguito il titolo di dottore di ricerca in Matematica nel 2019 presso l'Università degli Studi di Bari Aldo Moro discutendo una tesi dal titolo "Nonnegative Matrix Factorization for Knowledge Extraction from Biomedical and other real world data". Nello specifico lavora nel campo dell'analisi numerica e dei metodi di ottimizzazione di tipo low-rank per la fattorizzazione di grandi matrici di dati. E' stata assegnista presso il Dipartimento di Ingegneria Elettrica e Informatica e titolare di borsa di studio presso l'Istituto Tumori Giovanni Paolo II, dove ha potuto continuare la sua ricerca in ambito matematico con applicazioni a contesti reali come l'ingegneria e lo studio di tumori. Ha effettuato un periodo di ricerca in Belgio presso l'Université de Mons sotto la supervisione del Prof. Nicolas Gillis, uno dei massimi esperti mondiali nell'utilizzo delle tecniche di fattorizzazioni nonnegative. E' impegnata nel campo della divulgazione delle scienze matematiche con l'associazione di

promozione sociale Alumni Mathematica, della quale è presidente, ed è fermamente convinta che la matematica sia alla base di gran parte dei processi applicativi moderni e che la collaborazione e lo scambio di idee con esperti di altri settori possa aiutare a mettere questa disciplina al servizio di tutti.

Sessione 2: *Tecniche per la riduzione della dimensionalità e misure di similarità in Python*

Francesca Mazzia e Antonella Falini

Dipartimento di Informatica,

Università degli Studi di Bari Aldo Moro

`francesca.mazzia@uniba.it`, `antonella.falini@uniba.it`

Tramite tecniche per la riduzione della dimensionalità, basate sull'impiego di fattorizzazioni ortogonali, come ad esempio l'analisi delle componenti principali (PCA) e la decomposizione a valori singolari (SVD), sarà possibile eliminare eventuali ridondanze e quindi, gestire dei dataset con dimensione ridotta. Infine, l'utilizzo di *misure di similarità* o di *dissimilarità*, permetterà di quantificare la vicinanza o la correlazione dei dati osservati.

La teoria verrà affrontata mediante esercitazioni interattive usando il linguaggio Python.

CV breve: Antonella Falini dall'agosto 2019 è ricercatrice a tempo determinato presso il dipartimento di Informatica, dell'università "Aldo Moro" di Bari. Ha conseguito il PhD presso l'Istituto di Geometria Applicata dell'università Johannes Kepler di Linz (Austria), tramite una borsa Marie-Curie. Ha condotto ulteriori periodi di ricerca presso la Scuola di Computer Science di Cardiff (U.K.) e la Simplewere Ltd (U.K.). Ha effettuato un periodo di post-doc di tre anni presso l'Università di Firenze, tramite borse di ricerca INdAM. Il suo ambito di ricerca principale riguarda la modellistica e l'approssimazione numerica. Recentemente si occupa dello studio di nuovi modelli di clusterizzatori e di fattorizzazioni per il Big Data Analytics.

CV breve: Francesca Mazzia è professoressa ordinaria di Analisi Numerica presso il Dipartimento di informatica dell'Università degli studi di Bari Aldo Moro da gennaio 2018. Si è laureata in Scienze dell'Informazione nel 1989, dal 1990 ha svolto il ruolo di ricercatore e poi di professore associato presso la stessa Università. Nel 1997 ha svolto attività di ricerca presso il Cerfacs (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique) Toulouse, Francia, vincitrice di una borsa post-doc Marie-Curie. È autrice numerose pubblicazioni su riviste internazionali. La sua attività di ricerca riguarda principalmente i metodi numerici per la soluzione di equazioni differenziali ordinarie, sia ai valori iniziali che al contorno, attualmente si sta occupando di nuovi modelli di clusterizzatori e di fattorizzazioni approssimate per il BigData Analytics.

Sessione 3: *Introduzione al Clustering*

Claudia Angelini

Laboratorio di Statistica e Strumenti di Calcolo per la Bioinformatica
Istituto per le Applicazioni del Calcolo “Mauro Picone” - CNR, Napoli
`claudia.angelini@cnr.it`

In questa lezione verrà introdotto il problema del clustering, saranno presentati i principali metodi per il clustering gerarchico (di tipo aggregativo e divisivo), per il clustering partizionale (K-means e PAM) e verrà fornita una panoramica riguardo ulteriori strategie di clustering quali ad esempio gli approcci basati sulle misture. Successivamente, saranno illustrati i metodi per la scelta del numero di clusters e per la validazione di una clusterizzazione.

Durante le esercitazioni, le metodologie di clustering precedentemente descritte da un punto di vista matematico saranno illustrate attraverso alcuni esempi pratici svolti mediante R.

CV breve: Claudia Angelini si è laureata in Matematica nel 1994 presso l’Università degli Studi di Napoli “Federico II”, dove nel 2002 ha anche conseguito il dottorato di ricerca in Matematica Applicata ed Informatica. Dal 2001 lavora come ricercatore presso l’Istituto per le Applicazioni del Calcolo, CNR. Dal 2020 ricopre il ruolo di Dirigente di Ricerca. La sua attività principale è dedicata allo sviluppo di nuovi metodi statistici e di apprendimento automatico per l’analisi di dati complessi di grandi dimensioni. È stata coordinatrice scientifica dell’unità IAC-CNR in diversi progetti di ricerca e industriali. Attualmente è coordinatrice dell’unità di Biostatistica e Bioinformatica per la piattaforma tecnologica ADViSE. È stata co-autrice di oltre 70 articoli completi su riviste internazionali peer-reviewed ISI e di numerose altre pubblicazioni internazionali in atti di conferenze e capitoli di libri. Negli anni, ha supervisionato l’attività di ricerca di diversi dottorandi, studenti di master ed assegnisti di ricerca.

Sessione 4: *Aspetti di base relativi alla Regressione: teoria e applicazioni*

Antonio Di Crescenzo e Barbara Martinucci

Dipartimento di Matematica dell'Università degli Studi di Salerno

adicrescenzo@unisa.it, bmartinucci@unisa.it

La regressione é finalizzata ad individuare una relazione funzionale tra due o piú variabili osservabili, avendone a disposizione dei dati campionari. Nella prima parte sará presentato il modello di regressione lineare semplice e si illustreranno metodi per la stima dei coefficienti di regressione basati sulla tecnica dei minimi quadrati. Sará mostrato come tale metodo possa essere utilizzato anche nel caso di alcune equazioni di regressione di tipo non lineare. Inoltre, saranno discussi i principali strumenti per valutare la bontá dell'approssimazione ottenuta. Le metodologie presentate saranno poi illustrate attraverso esempi applicativi realizzati mediante R. Nella seconda parte sará affrontato il problema dei minimi quadrati pesati e saranno dati alcuni cenni alla regressioni polinomiale e alla regressione multipla.

CV breve: Antonio Di Crescenzo é professore ordinario di Probabilità e Statistica Matematica presso il Dipartimento di Matematica dell'Università degli Studi di Salerno, e membro del collegio dei docenti del Dottorato di Ricerca in Matematica, Fisica ed Applicazioni. È responsabile del Gruppo PRISMA (P**RO**bability In Statistics, Mathematics and Applications) dell'Unione Matematica Italiana. I suoi interessi di ricerca includono la teoria e la simulazione dei processi stocastici con applicazioni alla modellistica in biomatematica e ai sistemi di file d'attesa. Si dedica inoltre a problemi nell'ambito della teoria dell'affidabilità, e ad applicazioni in modellistica stocastica. È autore di numerose pubblicazioni in riviste internazionali, ed è membro del comitato editoriale di alcune riviste scientifiche, tra cui *Methodology and Computing in Applied Probability* (Springer Nature), *AIMS Mathematics* (AIMS Press), *Journal of Mathematics* (Hindawi), *Mathematics* (MDPI). È stato tutor per diverse tesi di dottorato di ricerca. Ha partecipato all'organizzazione di vari congressi internazionali e workshop.

CV breve: Barbara Martinucci è Professore Associato di Probabilità e Statistica Matematica presso il Dipartimento di Matematica dell'Università degli Studi di Salerno. Laureata con lode in Matematica nel 2000 presso l'Università di Napoli Federico II, ha ivi frequentato il Dottorato di Ricerca in Matematica Applicata e Informatica, conseguendo il titolo di Dottore di Ricerca nel 2005. Dal 01/10/2005 al 30/09/2008 è stata titolare di un assegno per la collaborazione ad attività di ricerca relativo al settore scientifico disciplinare MAT/06 presso l'Università degli Studi di Salerno. Nel periodo 01/11/2008 - 30/09/2020 è stata Ricercatore di Probabilità e Statistica Matematica presso l'Università degli Studi di Salerno. I suoi interessi di ricerca includono la teoria e la simulazione dei processi stocastici, con applicazioni in biomatematica. I contributi della sua attività di ricerca sono contenuti in circa 50 pubblicazioni in riviste scientifiche internazionali, atti di convegno e capitoli di libro. È stata membro del Comitato Organizzatore di diversi convegni e scuole internazionali. E' membro dell'Editorial Board di *Journal of Mathematics*

Sessione 5: *Introduzione alle reti neurali artificiali e reti convoluzionali per l'elaborazione delle immagini mediche*

Monica Bianchini e Simone Bonechi

Dipartimento di Ingegneria dell'Informazione dell'Università di Siena

nica@diism.unisi.it, simone.bonechi@unisi.it

Introduzione alle reti neurali feedforward

- La metafora neurobiologica;
- Reti neurali e soft-computing;
- Paradigmi di apprendimento;
- I modelli a singolo neurone: il Perceptron; l'algoritmo di addestramento;
- I modelli di neurone artificiale;
- Le reti feedforward e multistrato per classificazione e regressione;
- L'algoritmo di Backpropagation;
- Cenni a reti RBF e autoassociatori;
- Le reti convoluzionali;
- Gli strati convoluzionali e di pooling;
- Funzioni di attivazione e strati fully-connected;
- Loss function e ottimizzazione;
- Tecniche per sopperire alla mancanza di dati-

Introduzione ed esercitazione con la libreria di machine learning TensorFlow. La parte di esercitazione farà vedere come preparare i dati, definire una rete neurale, addestrarla (utilizzando tensorboard per monitorare le performance) e salvare il modello. Inoltre, verranno mostrati alcuni esempi di multi layer perceptron, rete convoluzionale, rete ricorrente e verrà spiegato come caricare un modello di rete pre-addestrato da utilizzare in un nuovo task. Infine verrà analizzato un caso pratico di utilizzo nell'ambito dell'analisi di lesioni cutanee.

CV breve: Monica Bianchini (<https://www.unisi.it/ugov/person/9664>) ha conseguito la laurea con lode in Matematica (1989) ed il dottorato in Ingegneria Informatica e dell'Automazione (1995) presso l'Università di Firenze. Dopo la laurea, ha partecipato per due anni ad un progetto congiunto Bull HN Italia-Dipartimento di Matematica (Univ. Firenze), per lo sviluppo di software parallelo per la risoluzione di equazioni differenziali. Dal 1992 al 1998 è stata studente di dottorato e, successivamente, borsista postdoc presso il Dipartimento di Sistemi e Informatica

dell'Università di Firenze. Dal 1999, svolge la propria attività presso il Dipartimento di Ingegneria dell'Informazione dell'Università di Siena, dove riveste attualmente il ruolo di professore associato (abilitazione al ruolo di prima fascia conseguita in data 14/04/2021).

L'attività di ricerca di M. Bianchini è focalizzata su machine learning, con particolare riferimento a reti neurali per dati strutturati, deep learning, teoria dell'approssimazione, information retrieval, bioinformatica ed elaborazione di immagini. M. Bianchini è autore di oltre cento pubblicazioni su rivista e atti di convegni internazionali (<http://scholar.google.it/citations?user=cKB0wzwAAA>) ed ha curato la pubblicazione di libri e di numeri tematici speciali per riviste del proprio settore di ricerca. Ha partecipato a numerosi progetti di ricerca su machine learning e pattern recognition, ed è stata coinvolta nell'organizzazione di eventi scientifici, quali il NATO Advanced Workshop on Limitations and Future Trends in Neural Computation (2001), l'VIII Convegno AI*IA (2002), il Convegno GIRPR 2012, l'International Symposium LOPSTR 2015 e l'ACM International Conference on Computing Frontiers (2017). M. Bianchini ha svolto attività editoriale come Associate Editor per IEEE Trans. on Neural Networks (2003-09), Neurocomputing (dal 2002) e Int. J. of Computers in Healthcare (dal 2010) ed è membro permanente dell'Editorial Board di IJCNN, ICANN, ICPR, ICPRAM, ESANN, ANNPR e KES.

CV breve: Simone Bonechi si è laureato in Information Engineering presso l'Università di Siena dove ha poi conseguito il PhD collaborando anche con l'Università di Copenaghen dove ha trascorso 3 mesi. Da febbraio 2021 è titolare di un assegno di ricerca presso il Dipartimento di Informatica dell'Università di Pisa in collaborazione con GSK. La sua attività di ricerca è focalizzata su deep learning e artificial intelligence, con particolare riferimento a analisi dati, computer vision ed elaborazione di immagini. Nello specifico ha studiato soluzioni per l'utilizzo di convolutional neural network in assenza di grandi quantità di dati e negli ultimi anni ha lavorato a vari progetti per lo sviluppo di strumenti di supporto alle decisioni in ambito biomedicale (es. analisi lesione cutanee, analisi fondo oculare, ...).

Seminario Aziendale

Data Science e Artificial Intelligence: come trasformare i dati in un vantaggio competitivo per le aziende

Alfredo Abrescia

Finconsgroup srl

Oggi i dati vengono generati e scambiati attraverso Internet senza sosta. E' di fondamentale importanza saper analizzare ed interpretare questi "grandi flussi di dati", ed estrarne valore che li trasformino in vantaggio competitivo per le aziende. Durante il seminario saranno mostrati i vantaggi della Data Science, i principali ambiti aziendali in cui è utilizzata e le principali tecnologie di Artificial Intelligence in grado di supportare il business aziendale.