UNIVERSITÀ DI PISA

COMPUTING METHODS FOR EXPERIMENTAL PHYSICS AND DATA ANALYSIS

ANDREA RIZZI

Academic year	2023/24
Course	FISICA
Code	360BB
Credits	9

Modules	Area	Type	Hours	Teacher(s)
COMPUTING METHODS FOR EXPERIMENTAL PHYSICS AND DATA ANALYSIS	FIS/01	LEZIONI	54	LUCA BALDINI GIANLUCA LAMANNA ALESSANDRA RETICO ANDREA RIZZI

Obiettivi di apprendimento

Conoscenze

Alla fine del corso lo studente avra' acquisito dimestichezza con i seguenti strumenti per il calcolo scientifico e l'analisi dati:

- comandi avanzati shell unix per automazione di operazioni
- linguaggio python e principali moduli di sistema
- librerie python per calcolo scientifico
- strumenti per la gestione e documentazione del software
- tecniche e librerie per calcolo parallelo e su GPU
- conoscenze dei concetti fondamentali in machine learning e per sviluppo di reti neurali
- reti feed-forward, reti convolutive e reti ricorsive
- reti generative (GAN e autoencoders) e graph networks
- strumenti specifici per applicazioni in fisica delle particelle e/o fisica medica (*)

(*) non disponibile per la versione da 6 CFU del corso

Modalità di verifica delle conoscenze

Sara' effettuata una verifica delle conoscenze acquisite attraverso domande specifiche in sede di esame orale partendo da un progetto di calcolo scientifico realizzato e presentato dagli studenti

Capacità

Al termine del corso lo studente sapra':

- Automatizzare le operazioni di calcolo nelle analisi dati
- Scrivere programmi per calcolo scientifico in linguaggio python
- Sviluppare software in ambienti altamente collaborativi
- Scrivere programmi che sfruttino architetture di calcolo parallelo
- Progettare reti neurali per analisi dati
- Sviluppare reti neurali per calcolo scientifico
- Sviluppare progetti di analisi di dati di fisica delle particelle o di fisica medica (*)

(*) non disponibile per la versione da 6 CFU del corso

Modalità di verifica delle capacità
Le capacita' saranno verificate richiedendo allo studente lo sviluppo di un progetto di software per analisi dati e/o l'analisi di un set di dati attraverso le tecniche apprese nel corso.

Comportamenti
Il corso ha una parte interamente dedicata alle best practice nello sviluppo di codice specialmente in un'ottica di collaborazioni scientifiche. Il corso insegnera' come documentare e condividere gli strumenti di calcolo scientifico sviluppati.

Modalità di verifica dei comportamenti
La verifica delle capacita' acquisite sara' fatta sia richiedendo l'utilizzo degli strumenti di collaborazione per la presentazione degli elaborati, sia verificando l'attenziene nella parte di documentazione dei progetti presentati.

Prerequisiti (conoscenze iniziali)
E' necessaria la conoscenza minima di funzionamento di un computer.
Sarebbe inoltre consigliabile possedere conoscenze, anche rudimentali, di programmazione in altri linguaggi (es. C)

Corequisiti

Prerequisiti per studi successivi

Indicazioni metodologiche
Il corso sara' composto da lezioni frontali teoriche e da lezioni frontali in cui si affronta un problema mostrando come esercizio lo sviluppo di un particolare algoritmo, la realizzazione di un programma o l'analisi di un set di dati con una particolare tecnica. Inoltre saranno fatte esercitazioni in cui gli studenti dovranno sviluppare sul proprio computer alcuni gli esercizi proposti.

Programma (contenuti dell'insegnamento)
Automazione delle operazioni (shell), Version control, introduzione al python, algoritmi e complessita, contenitori python, algoritmi standard, iteratori, funzioni lambda, classi e oggetti, overload, scipy/numpy e gestione array, matplotlib, tecniche classiche di ML, issue tracking e ambienti collaborativi, unit test, gestione delle memoria e profiling. Performance CPU/GPU, I/O verso varie periferiche e latenza, evoluzione

dell'hardware per calcolo scientifico, programmazione su GPU, vettorizzazione, reti neurali, CNN, RNN, keras e tensorflow, metodi di training, autoencoders, GANs, GraphNetworks. Introduzione al C++, STL, C++11 and C++14, ROOT toolkit, PyROOT, Medical data processing and feature extraction, handling standard-format medical data (DICOM), data anonymization, visualization, Deriving features form images, Image segmentation,Data quality control pipelines, outlier removal, dimensionality reduction, Data analysis and classification, figures of merit, cross-validation, permutation test, Machine-learning and deep-learning tools for segmentation and classification (specific platforms for medical data),Data augmentation, transfer learning, retrieving localization information

Bibliografia e materiale didattico

https://www.deeplearningbook.org/
https://www.learncpp.com/cpp-tutorial/introduction-to-cplusplus/
The Hundred-Page Machine Learning Book, A.Burkov
Graph Representation Learning, William L. Hamilton

Indicazioni per non frequentanti

Modalità d'esame

Presentazione di un elaborato su un progetto concordato con i docenti ed esame orale di discussione dell'elaborato e verifica della conoscenza teorica degli strumenti utilizzati nel progetto, nonche' delle scelte operate nella realizzazione del progetto

Stage e tirocini

Altri riferimenti web

Note

Ultimo aggiornamento 31/08/2023 15:23