AREA SCIENCE PARK

PROFILO B

Traccia n. 1
Il candidato discuta in maniera sintetica due dei seguenti argomenti, indicando esplicitamente la scelta all'inizio del proprio elaborato:

A) Il candidato descriva brevemente il metodo di regressione logistica binaria e indichi un possibile ambito di applicazione. Si assuma che i dati di input siano costituiti da vettori di dimensione M. Il candidato risponda alle seguenti domande:

- Qual è la funzione di "loss" più utilizzata, come viene definita e qual è la sua interpretazione statistica/probabilistica?
- La regressione logistica binaria può essere descritta in termini di reti neurali: qual è la rete neurale equivalente e quali sono i suoi elementi architetturali?
- Quali sono le limitazioni di espressività della regressione logistica binaria e quali strategie possono essere adottate per superare queste limitazioni?

B) Il candidato descriva brevemente cosa sia una Rete Neurale Convoluzionale (CNN), spiegare per quali tipi di problemi può essere utilizzata e delinearne gli elementi architetturali caratteristici. Il candidato risponda alle seguenti domande:

- Quali sono le principali proprietà dell'operazione di convoluzione (così come intesa in questo contesto specifico) e quali sono le motivazioni per la sua introduzione?
- Qual è il nuovo elemento architetturale introdotto per la prima volta nell'architettura della rete convoluzionale ResNet, quali vantaggi offre e quale problema affronta?

C) Si consideri un laboratorio dotato di strumenti scientifici in grado di generare grandi volumi di dati. I dati grezzi devono essere trasferiti e salvati automaticamente in un centro di calcolo. Una volta salvati con successo, deve avviarsi automaticamente una pipeline di analisi dei dati. Inoltre, i dati e gli strumenti computazionali correlati devono essere resi accessibili ai ricercatori tramite una piattaforma cloud che supporti l'uso di strumenti open-source per l'analisi, il machine learning e il deep learning. L'obiettivo è garantire che la ricerca sia il più possibile riproducibile. Il candidato delinei gli elementi essenziali di un'infrastruttura di ricerca di questo tipo. Inoltre, il candidato affronti brevemente almeno due dei seguenti argomenti:

- Gestione dei dati secondo i principi FAIR
- Strumenti di gestione dei workflow
- Definizione di schemi di metadati

Traccia n. 1
The candidate should discuss, in a concise and effective manner, two of the following topics, explicitly indicating their choice at the beginning of their work:

AREA SCIENCE PARK

A) The Candidate should briefly describe the binary logistic regression method and indicate a possible area of application. Assume that the input data consists of vectors of dimension M. The Candidate should also answer the following questions:

- What is the most used "loss" function, how is it defined, and what is its statistical/probabilistic interpretation?
- Binary logistic regression can be described in terms of neural networks: what is the equivalent neural network, and what are its architectural elements?
- What are the expressiveness limitations of binary logistic regression, and what are possible strategies to overcome these limitations?

B) The Candidate should briefly describe what a Convolutional Neural Network (CNN) is, explaining what types of problems it can be used for and outlining its characteristic architectural elements. The Candidate should answer the following questions:

- What are the main properties of the convolution operation (as understood in this specific context), and what are the motivations for its introduction?
- What is the new architectural element introduced for the first time in the ResNet convolutional network architecture, what are its advantages, and what problem does it address?

C) Consider a laboratory equipped with high-throughput scientific instruments capable of generating large volumes of data. The raw data must be automatically transferred and stored in a computing center. Once the data has been successfully saved, a data analysis pipeline should automatically start. Additionally, the data and related computational tools must be made accessible to researchers via a cloud platform that supports the use of open-source tools for analysis, machine learning, and deep learning. The goal is to ensure that the research is as reproducible as possible. The Candidate should outline the essential elements of such a research infrastructure.

- Additionally, the Candidate should briefly address at least two of the following topics:
- FAIR Data Management
- Workflow management tools
- Definition of metadata schemas