

Avviso di Seminario

Lunedì 20 maggio 2024, ore 11.00

Aula D – Palazzo Fortuna

Dimensionality reduction and clustering based on the data intrinsic dimension

Antonietta Mira

Università della Svizzera italiana and Insubria University

Abstract: Even if they are defined on a space with a large dimension, data points usually lie onto manifolds with a much smaller intrinsic dimension (ID). How many variables are necessary to describe a real-world data set without significant information loss? What is the appropriate scale at which one should analyze and visualize the data? These two questions, which are often considered unrelated, are actually strongly entangled, and can be addressed within a unified framework. We introduce an approach in which the optimal number of variables and the optimal scale are determined self-consistently, recognizing and bypassing the scale at which the data are affected by noise. To this aim we estimate the data intrinsic dimension (ID) in an adaptive way. Sometimes within the same dataset it is possible to identify more than one ID meaning that different subsets of the data points lie onto manifolds with different IDs. Identifying these manifold provides a clustering of the data and in many real world application a simple topological feature like the ID allows to uncover a rich data structure thanks to manifold related clustering.

Webinar:

https://teams.microsoft.com/l/meetup-join/19%3ameeting_OTIoZTBlNmItNzIwNyooNDM2LWIwMTUtZDgwMjExYThlYzYo%40thread.v2/0?context=%7b%22Tid%22%3a%22baeefbc8-3c8b-4382-9126-e86bfef46ce6%22%2c%22Oid%22%3a%22fb8be958-7031-415b-bac3-d4c03fcf93fc%22%7d