CALCOLO SCIENTIFICO E MODELLI MATEMATICI ALLA RICERCA DELLE COSE NASCOSTE ATTRAVERSO LE COSE MANIFESTE

Title: Data-driven non-linear approximation methods for dimensionality reduction of hyperspectral images

Authors: Vittoria Bruni, Giuseppina Monteverde, Domenico Vitulano

Affiliation: Department of Basic and Applied Sciences for Engineering - Sapienza Rome University

Via Antonio Scarpa 16, 00161 Rome, Italy

Abstract: Hyperspectral imaging (HSI) is a non-invasive and non-contact technology that can identify the composition of an object based on tens of hundreds of wavelengths reflected from the object and visualize their spatial distribution. Thanks to their high spectral resolution, hyperspectral images (HI) give useful information about the object, that human eye is not able to see. On the other hand, HI are high dimensional data that require significant computational sources especially when embedded in Machine Learning (ML) and Deep Learning (DL) architectures. Dimensionality reduction problem in HSI can be handled through two different approaches: i) selecting significant spectral features in a transformed domain; ii) selecting relevant spectral bands from the original spectrum. In the former case, non-linear approximation methods in a proper expansion basis are required; in the latter, non-uniform sampling methods are better suited. Two different approaches for dimensionality reduction of HI are proposed for data classification purposes with the aim of exploiting the advantages of ML and DL in providing high classification rates, while reducing the computational load and time. The first method projects the data in a proper domain and automatically selects relevant components in the new space based on the entropic normalized information distance. The latter measures the complexity of the projected data in terms of coding length. The second method applies a non-uniform sampling of the single spectra by exploiting the properties of the wavelet transform in representing signal non-stationarity. It is an adaptive and automatic method where the number of bands is not pre-defined but automatically given. The properties of the two methods are discussed and their advantages when embedded in ML/DL based data classification are evaluated through extensive comparative studies using benchmarking datasets.