Methods for Feature Learning and Extraction

Feature learning is widely regarded as a fundamental basis for many modern machine learning algorithms, specifically those classified under unsupervised learning. Its applications span a wide variety of fields, ranging from determining the connectedness of airports to classifying an apartment’s location based solely on a few properties. Researchers have focused extensively on image recognition and classification. It is intriguing that the human brain can easily process and interpret images, yet even extremely powerful computers struggle to do the same. It is only logical then that researchers have turned to the human brain for inspiration and have attempted to mimic the visual cortex’s operations in code. Several methods and algorithms have developed over the years with some remarkable success.

Sparse coding is one algorithm that draws heavy influence from biology. It makes the logical assumption that the brain evolved to cope with images by somehow decoding them into information that could easily be interpreted. In the process, certain information is discarded to remove redundancies in the data. Sparse coding is designed to remove redundancies in the input to create a code matrix consisting of statistically independent basis vectors. This reduces the problem to the seemingly simple equation where is the coded vector, is an invertible weight matrix, and is the image.

While such an equation is usually sufficient for a task such as image compression, it does not take many of the non-linear features of image classification into account. An alternative method that is more suitable is to create a set of synthesis function along with associated amplitudes. In mathematical terms, this equates to:

As evidenced from the above equation, the goal is to now determine a set of basis functions that can accurately represent the basic structure of the images. One set of solutions that are rather interesting are those that are overcomplete, or where the number of basis functions exceeds the dimensionality of the input.