Non-negative Matrix and Tensor Factorization

Introduction

Many of the most descriptive features of speech are described by energy; for example, formants are peaks and the fundamental frequency is visible as a comb-structure in the power spectrum. A basic property of such features is that they are positive-valued. Negative values in energy are not physically realizable. However, most signal processing methods are applicable only for real-valued variables and inclusion of a non-negative constraints is cumbersome.

Non-negative matrix factorization (NMF or NNMF) and its tensor-valued counterparts is a family of methods which explicitly assumes that the input variables are *non-negative*, that is, they are by definition applicable to energy-signals. In some sense, NMF methods are an extension of principal component analsys (PCA) -type and other subspace methods to positive-valued signals.

Model definition

Specifically, suppose that the power (or magnitude) spectrum of one window of a speech signal is represented as a *Nx1* vector v_k, and furthermore we arrange the *K* windows into an *NxK* matrix *V*. The signal model we use is then

\[V \approx WH, \]

where *W* is the *NxM* weight matrix, *H* is the *MxK* model matrix and the scalar *M* is the model order.

The idea is that *H* is a fixed matrix corresponding to our model of the signal, viz. the source model. It describes typical types features of the data. With the weights *W*, we interpolate between the columns of *H*. In some sense, this is then a generalization of a codebook (see vector quantization), but such that we interpolate between codevectors. In addition, we require that all elements of *W* and *H* are non-negative, such that we ensure that *V* is also non-negative.

Since the model order *K* is chosen to be smaller than either *N* or *K*, this mapping is generally an approximation. The model thus tries to catch *the relevant features of the input signal with a low number of parameters*.

The model is generally optimized by

\[\min_{W,H} \| V - WH \|_F\qquad\text{such that}\qquad W,H\geq 0. \]

Here the norm refers to the Frobenius norm, which is defined as the square root sum of squared elements. We do not have analytic solutions to the above optimization problem, but we can solve it by numerical methods, which are included in typical software libraries.

Application

A typical use of NMF type algorithms is source separation, where we find the solution of the above optimization problem and then identify those dimensions of *H* which corresponds to the different sources. By retaining only those dimensions of *W* which correspond to the desired source, we can thus extract the desired source signal from their mixture with the interfering other sources. For example, we might want to extract a speech signal corrupted by noise by extracting the dimensions corresponding to speech and removing those dimensions which correspond to noise.

Note however that NMF-type methods extract only the power (or magnitude) spectrum of the desired signal. In contrast, usually the input signal is a time-frequency representation which has also a phase-component. After application of NMF-estimation, we therefore need also an estimate of the phase-component of the signal. Such methods will be discussed in the speech enhancement chapter of this document.

For more information, see the Wikipedia article: Non-negative matrix factorization.