NON-NEGATIVE MATRIX FACTORIZATION AND APPLICATIONS

PIPER MORRIS, BRANDON BONIFAS-REYES & RAD MALLARI[1]

May 27, 2022

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Non-negative matrix factorization, often referred to as NMF, is one of many unsupervised learning algorithms used to extract significant figures from large sets of data. It is a relatively new process that has become an important tool in dimensionality reduction, and is comparable to other factorization techniques such as principal component analysis (PCA) and vector quantification (VQ). What makes NMF unique is its constraint on non-negative elements resulting in an often preferable outcome. NMF has naturally garnered popularity in industries such as machine learning because the algorithms efficiently extract sparse data and provide respective factors that one can easily interpret.

[1] *Department of Mathematics, University of California, Santa Barbara*

1 INTRODUCTION

A statement requiring citation [1].

Given an $n \times m$ non-negative matrix A, NMF aims to express A as two similarly non-negative matrices of smaller dimensions, W and H. Note that when referring to a non-negative matrix, we are discussing a matrix in which all its entries are real and greater than or equal to zero. W is an $m \times r$ matrix consisting of k basis elements of A and H is a $r \times n$ matrix consisting of the coefficients, or weights, related to the entries of W. We denote some $r > 0$ as the inner dimension of W and H and the smallest such instance where r holds is called the *factorization rank*. While there are several ways to compute this rank to be discussed later, it is important to note that r must be at least the size of the rank of A. Because our resulting NMF is a nonnegative approximation, W, H (holding the same properties) may not be unique either. Nonetheless, we utilize methods for measuring the error of our approximations.

2 COST FUNCTIONS

As mentioned above, the problem of complexity with NMF arises out of the fact that its algorithms attempt to replicate the original matrix, namely A. Thus, the product of WH will be a representation of A and not an exact replica. As a result, we implement various methods to go about solving an NMF problem. One of the most common ways to execute this is by using a cost function of which there are several. They share a common goal in that they attempt to minimize the error between our original matrix and the product of our factorization. One useful method is accomplished by minimizing the square of the Euclidean distance. This is also often referred to as minimizing the Frobenius norm as it relates specifically to matrices.

$$\|A - WH\|_F^2 = \sum_{ij} (A - WH)_{ij}$$

This measure is often chosen because it is not as expensive as other cost functions in terms of calculation. It also has a key property in that it remains invariant under rotations/orthogonal transformations. Second, it accounts for the presence of Gaussian noise. (It is also recognized as the Euclidean norm because it compiles all the rows/columns of a matrix and concatenates them to produce one vector.) <- Delete?? Here, one uses iterative update rules to exchange a randomized WH matrix with one that has a smaller error. We continue to do this until the error meets a predetermined requirement.

An alternative norm used in the process of decomposition is the Kullback Liebler divergence:

$$D_{KL}(A|WH) := \sum_{i=1}^{m}\sum_{j=1}^{n}((WH)_{ij} - A_{ij}\log(WH)_{ij}) + \sum_{i=1}^{m}\sum_{j=1}^{n}(A_{ij}\log A_{ij} - A_{ij})$$

3 MULTIPLICATIVE UPDATE RULE

A popular approach to NMF problems is the multiplicative update rule created by Lee and Seung. The multiplicative update rule became popularized because of its simplicity; however, it's criticized for its slow convergence and because there is no guarantee of convergence to a stationary point which is necessary to find a local minimum.

Formula is given by:

$$W \leftarrow W \cdot \frac{(VH^{T})}{(WHH^{T})}$$

$$H \leftarrow H \cdot \frac{(W^{T}V)}{W^{T}WH}$$

4 APPLICATIONS

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

4.1 Image Processing

NMF often begins its canonical relationship with the real world in the field of image processing. We take a simple image of a face with p pixels, and reduce the dimensions of the data to a single vector such that the ith entry represents the ith pixel. Then we can allow the rows of our matrix A to represent the p pixels and n columns respectively output one image. The process of NMF, as we have defined it, will produce W and H to be multiplied, where the columns of W embody the basis images and H gives instruction on how to sum up the aforementioned images. All this to reconstruct an approximation of the originally provided face.

4.2 Text Mining

We utilize NMF in the field of topic recovery/document classification. Let each column of A correspond to a text and each row relate to a keyword. Construct every (i,j)-th entry in such a way where it represents every time a given word (i) comes up in a document (j). This is a type of quality of term frequency construction in such a model, namely the bag-of-words model, where each document being analyzed has its respective set of words. The ordering of words here is not accounted for. The concept of sparseness appears once more in A due to the nature of texts/documents which only use a portion of the dictionary. In any case, NMF provides a decomposition of rank r as follows:

$$A(i,j) \approx \sum_{k=1}^{r} W(i,k) \cdot \sum H(k,j)$$

Note that in the above A is our j-th document, W our k-th topic, and H conveys the relevance of a topic in a given document.

4.3 Facial Recognition

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

REFERENCES

[1] A. J. Figueredo and P. S. A. Wolf. Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330, 2009.