

# Non-negative Matrix Factorization for Medical Imaging

Miguel Atencia<sup>1</sup> and Ruxandra Stoean<sup>2</sup>

1- Universidad de Málaga - Departamento de Matemática Aplicada  
Escuela de Ingenierías Industriales, 29071-Málaga - Spain

2- University of Craiova - Department of Computer Science  
Faculty of Sciences, 200585-Craiova - Romania

**Abstract.** A non-negative matrix factorization approach to dimensionality reduction is proposed to aid classification of images. The original images can be stored as lower-dimensional columns of a matrix that hold degrees of belonging to feature components, so they can be used in the training phase of the classification at lower runtime and without loss in accuracy. The extracted features can be visually examined and images reconstructed with limited error. The proof of concept is performed on a benchmark of handwritten digits, followed by the application to histopathological colorectal cancer slides. Results are encouraging, though dealing with real-world medical data raises a number of issues.

## 1 Introduction

A non-negative matrix factorization (NMF) method for learning reduced feature sets for image classification is introduced in the current study. The approach is important for real-world learning tasks, where images have a massive dimension than may hinder the classification process. On the other hand, distinguishing between the redundant knowledge and the discriminative indicators is a critical aspect of medical images, whose supervised labelling is often noisy. Nevertheless, the task is worthwhile, given the decisive information that classification can provide to diagnosis support.

In this context, the aim of the proposed approach is twofold. Firstly, it will be able to discover the inside feature components that can be further visually inspected. Secondly, the reduced matrix obtained from the factorization can be subsequently used in the training part of the classification instead of the larger initial data matrix, with the given labels. Additional test examples can be encoded and classified, allowing for efficient prediction.

A first test case demonstrates the effectiveness of the methodology in benchmark conditions. The second application of the NMF is performed on the challenging task of histopathological slide interpretation in search of cancerous patterns for the colorectal disease.

The structure of the paper is set as follows. The NMF technique for image reinterpretation is put forward in section 2. The proof of concept on handwritten digit data is outlined in subsection 3.1, whereas the application to the medical task is described in subsection 3.2. Conclusions are drawn in the enclosing section of the paper.

## 2 Non-negative Matrix Factorization for Image Encoding

The current study proposes a NMF methodology to model dimensionality reduction and subsequent classification for image data sets. The approach follows several steps. Non-negative factorization of an  $n \times m$ -sized matrix  $Y \geq 0$ , where the inequality is considered component-wise, yields two factors  $W, H$ , both also non-negative, so that the approximation error  $\|Y - WH\|$  is minimized. Several choices for the matrix norm are possible, but in this work we restrict ourselves to the most usual Frobenius norm. If the sizes of  $W$  and  $H$  are  $n \times r$  and  $r \times m$  respectively, the algorithm is meaningful when  $r \ll \min(n, m)$  and, as such, it has been related to Principal Components Analysis and Low Rank Approximation. For notation, background and examples, see [1] and references therein.

NMF has been applied to several tasks related to feature selection and dimensionality reduction, including text mining and source separation. In the particular context of image processing,  $p \times q$  resolution images are first reshaped to  $n = pq$ -dimensional vectors, so that each image forms a column of an  $n \times m$  matrix  $Y$ , where  $m$  is the number of images available. The matrix is subsequently decomposed as  $Y \approx WH$ , where the key parameter  $r$  is set manually. In this work we implement the Hierarchical Alternating Least Squares (HALS) algorithm, which solves the unconstrained least squares problem alternatively for one column of  $W$  and one row of  $H$  at a time, followed by projection to ensure non-negativity. Initialization, which is known to be a critical aspect of the algorithm, is based upon Singular Value Decomposition [2]. Also note that the algorithm allows for easy scaling or normalizing the columns of  $W$ : for each column  $W_{.j}$ , compute  $\alpha_j = \max(W_{.j})$  and replace  $W_{.j} \leftarrow \frac{W_{.j}}{\alpha_j}$ , as well as the rows of  $H$  by  $H_{.j} = \alpha_j H_{.j}$ , so the product  $WH$  remains constant, and each column of  $W$  is distributed in the interval  $(0, 1)$ .

Since  $W, H$  are non-negative, if the columns of  $W$  are scaled to the same magnitude as those of  $Y$ , they can be interpreted as *features* or cluster centres within Principal Components Analysis and Self-Organizing Maps [3]. In turn, each column  $H_{.i}$  constitutes an *encoded* or compressed version of the corresponding original image  $Y_{.i}$ , which could be later recovered by  $\hat{Y}_{.i} = WH_{.i}$ , and the encoding error can be assessed by  $\|Y_{.i} - \hat{Y}_{.i}\|$ .

In this work we propose the inclusion of the NMF algorithm as a previous step to the application of a classification method, which is expected to have an enhanced performance and reduced computational cost due to the much lower dimensionality of the trained vectors  $H_{.i}$ , in contrast to the original images  $Y_{.i}$ . Thus NMF will play a similar role to that of autoencoders in Deep Learning architectures [4], with the added benefit of easier interpretation of results as feature vectors. The  $W$  matrix is constructed as a result of the NMF algorithm, which can be considered a sort of *training*. Then, a subsequent *test* image  $y$  should be encoded into a compressed vector  $h$ , in such a way that  $y \approx Wh$ . The compression can be achieved by the Moore-Penrose pseudoinverse  $W^+ = (W^T W)^{-1} W^T$  and the low-dimensional code is obtained by  $h = W^+ y$ , as suggested by [5] in the context of source separation.

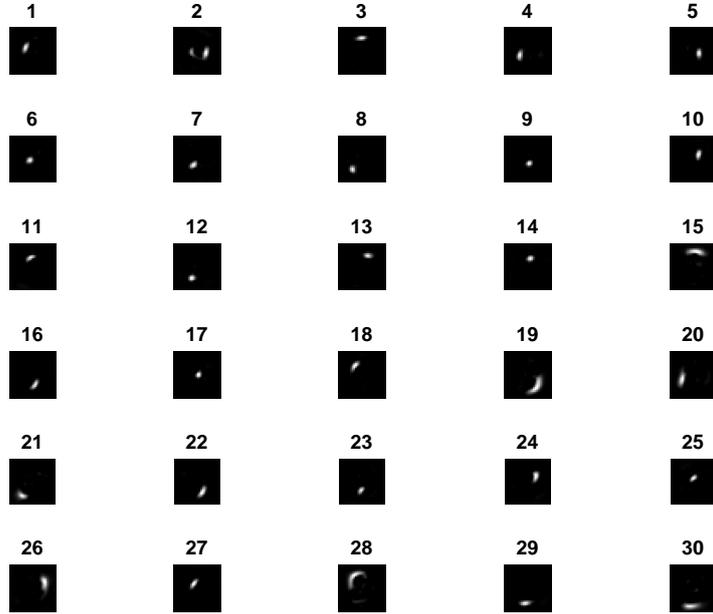


Fig. 1: Features or basis images in the resulting matrix  $W$  provided by NMF with  $r = 30$  on the handwritten digit data.

### 3 Experimental Results

The application of NMF to image data is expected to provide dimensionality reduction and subsequent faster learning of the classifier.

#### 3.1 Experiment 1: Proof of Concept on Handwritten Digits

The first experiment assesses the NMF algorithm by considering the benchmark synthetic handwritten digit data set provided by Matlab, which consists of training and test sets containing 5000 28x28 resolution sample images each. The task is then to categorize the 784-dimensional vectors into the 10 classes.

##### 3.1.1 Encoding

The key hyperparameter  $r$  is manually set as  $r = 30$ , which triggered the best performance within the experimental trials. The corresponding features appearing in the columns of  $W$  are shown in Figure 1. For the training set, the encoded images are directly recovered by  $\hat{y} = W h$  from the columns of  $H$  computed in the factorization. For the test set, the encoded version  $h = W^+ y$  is first computed, and decoded again by  $\hat{y} = W h$ . The encoding error for 5 random columns of the test set are depicted in Figure 2, where it can be seen that the reduced versions look similar to the original ones with a visually assessable small error.

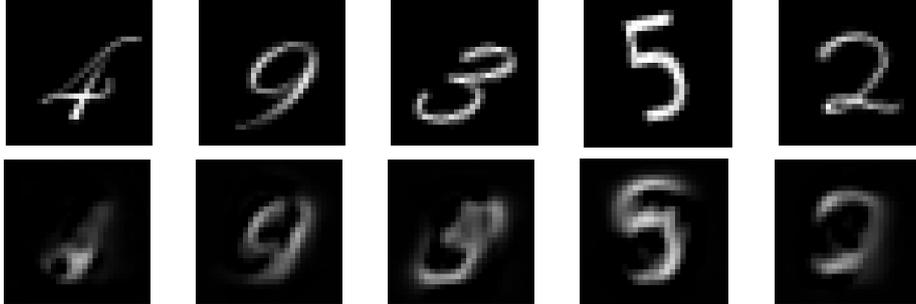


Fig. 2: Five handwritten digit test samples in original (top row) and encoded/decoded version (bottom row). Mean error: 2.46, 2.48, 3.23, 4.01, 2.36.

Table 1: Results of shallow classifiers on the original and reduced digit data.

Method	Configuration	Acc. (%)	Time (sec.)
<b>Initial data</b>			
SVM	linear, 1a1	90.96	137
MLP	1 hidden layer (30), 100 maxit, SCG	86.88	321
RF	5 variables at split, 30 trees	97.12	48
<b>Reduced H data</b>			
SVM	radial, 1a1	97.5	17
MLP	1 hidden layer (100), 300 maxit, SCG	95.98	191
RF	5 variables at each split, 30 trees	97.56	2

### 3.1.2 Learning

The next step is to use the labels and try to learn the image-label correspondence by means of a classifier. Three state-of-the-art shallow approaches are implemented in  $R$ , namely Support Vector Machines (SVMs), Multilayer Perceptron (MLP) and Random Forests (RF). These choices allow to assess the encoding ability before a variety of classification paradigms: statistical learning theory, connectionism, and ensemble learning.

Table 1 presents the results of the three employed shallow classifiers on the initial collection of digit images (with the training and test pre-decided splits) and on the reduced set given by  $H$  and the accordingly encoded  $H_{\text{test}}$ . The approaches were implemented in  $R$  with the corresponding packages *e1071*, *RSNNS* and *randomForest*. The parameter configurations for the three approaches, which are the result of manual tuning, are also outlined within Table 1. The NN architecture for the initial data set was chosen as a trade-off between time and accuracy. The accuracy given by the RF is the result of 30 repeated runs of the algorithm, given the stochastic nature of the learner.

There are several conclusions that can be drawn from examining the accuracy-time results alongside the necessary parametrization requirements in Table 1:

- The RF improves runtime when training on the compressed data  $H$ .
- For the MLP, the encoding allows for a larger architecture to be trained.
- The SVM, when using NMF, is much faster and reaches the RF accuracy.

### 3.2 Experiment 2: Real-world Scenario of Histopathological Images

The second experiment investigates the potential of the NMF on a more difficult as well as highly informative image data from the medical field. The problem tackles automatic cancer grading from histopathological digitalized slides for colorectal tissues [6]. The pathologist is supposed to establish the cancer stage by measuring the variation of the size, shape and texture of the glands and nuclei [7]. There are 96 G1 (grade 1) records, 99 G2 (grade 2) and 100 G3 (grade 3) images. The collection comes from the Emergency Hospital of Craiova, Romania <sup>1</sup>. The original slides are at resolution 800x600 and also contain benign instances, obtained by cutting healthy portions from tissue images of ill patients at the indication of the pathologist [8], since biopsy is usually performed only when there is suspicion of cancer presence.

For this experiment, the task is formulated as binary classification, i.e. to distinguish between G1 (cancer in an incipient stage) and G23 (advanced invasion). This discrimination is commonly necessary for prompt assessment by the physician, in order to decide subsequent therapy. For a direct processing of the images within the NMF and subsequent learning, they are compressed at 25%, i.e. having 200x150 pixels, and transformed to greyscale, thus resulting in 30000-dimensional input vectors. In order to account for the stochastic nature of classifiers, 10 splits of repeated random sub-sampling cross-validation with 70% training (250 samples in  $Y$ ) - 30% test (107 examples in  $Y_{\text{test}}$ ) are performed. Note that classification by a shallow approach of these initial samples can either be restricted by memory limits or be intractable in terms of computational time, depending on the implementation platform.

The NMF is applied to the training set with  $r = 100$ , which is close to the maximum imposed by the restriction  $r \ll \min(n, m)$ . The generated matrix  $H$  is then used for learning. The test set is obtained again by the encoding  $H_{\text{test}} = W^+ Y_{\text{test}}$ , as mentioned before in subsection 3.1.2.

The mean reached accuracy over the 10 trials is of 70%. The relatively high number of errors is probably due to the loss of information resulting from encoding 30000 pixels into only 100 features through the NMF. Additionally, histopathological slides are originally stained for a better visualization by the pathologist. The conversion to greyscale has certainly lost other important features regarding colour display between regions.

---

<sup>1</sup><https://doi.org/10.6084/m9.figshare.4508672.v1>

## 4 Conclusions

The paper tailored a NMF methodology to dimensionality reduction for image learning tasks. The approach was first validated on a benchmark collection of handwritten digits. The features were visually inspected and the encoded image versions were seen to share a great similarity with the initial ones. Learning only the reduced  $H$  matrix resulting from the NMF led to a test classification accuracy into the 10 categories of 97.5% in at least 20 times less runtime.

The second test instance, the histopathological data classification into less and greatly invasive colorectal cancer, was a preliminary study of the application of NMF as an image encoder for a real-world medical image task. The obtained results are acceptable, in the light of a drastic dimensionality reduction of the data from 30000 to 100 and the use of a greyscale transformation of otherwise stained images. These results highlight a critical issue when dealing with medical images: labelling is time-consuming and subjective, resulting in scarce and noisy instances for supervised classification. From a mathematical viewpoint, the data matrix has more rows than columns, thus limiting the number of basis images that can be obtained. The next obvious step is to artificially increase the number of slides, e.g. by partitioning a full 800x600 image into quarters. This will allow to increase the value of  $r$  within NMF resulting in a larger number of features, which should eventually lead to more accurate learning. Secondly, the use of the original RGB condition will probably comprise further important features in the discrimination process. Dealing with these *three-dimensional* matrices could be achieved by techniques of *tensor* factorization. Finally, the rank  $r$  of the factorization should be optimally set through a parameter tuning heuristic, independent of the size of the vectors under consideration for each problem.

## References

- [1] Nicolas Gillis. The Why and How of Nonnegative Matrix Factorization. In *Regularization, Optimization, Kernels, and Support Vector Machines*, pages 257–291. 2014.
- [2] C. Boutsidis and E. Gallopoulos. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362, April 2008.
- [3] Imola Fodor. A survey of dimension reduction techniques. Technical report, 2002.
- [4] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.
- [5] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-Ichi Amari. *Nonnegative Matrix and Tensor Factorizations*. John Wiley & Sons, Ltd, Chichester, UK, 2009.
- [6] Stefan Postavaru, Ruxandra Stoean, Catalin Stoean, and Gonzalo Joya Caparros. *Adaptation of Deep Convolutional Neural Networks for Cancer Grading from Histopathological Images*, pages 38–49. Springer International Publishing, Cham, 2017.
- [7] M.N. Gurcan, L.E. Boucheron, A. Can, A. Madabhushi, N.M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.
- [8] Olcay Sertel, Jun Kong, Umit V. Catalyurek, Gerard Lozanski, Joel H. Saltz, and Metin N. Gurcan. Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. *Journal of Signal Processing Systems*, 55(1):169–183, 2008.