

NMF (Non-negative Matrix Factorization)

NMF is a model that factors high-dimensional vectors into a low-dimensionality representation.

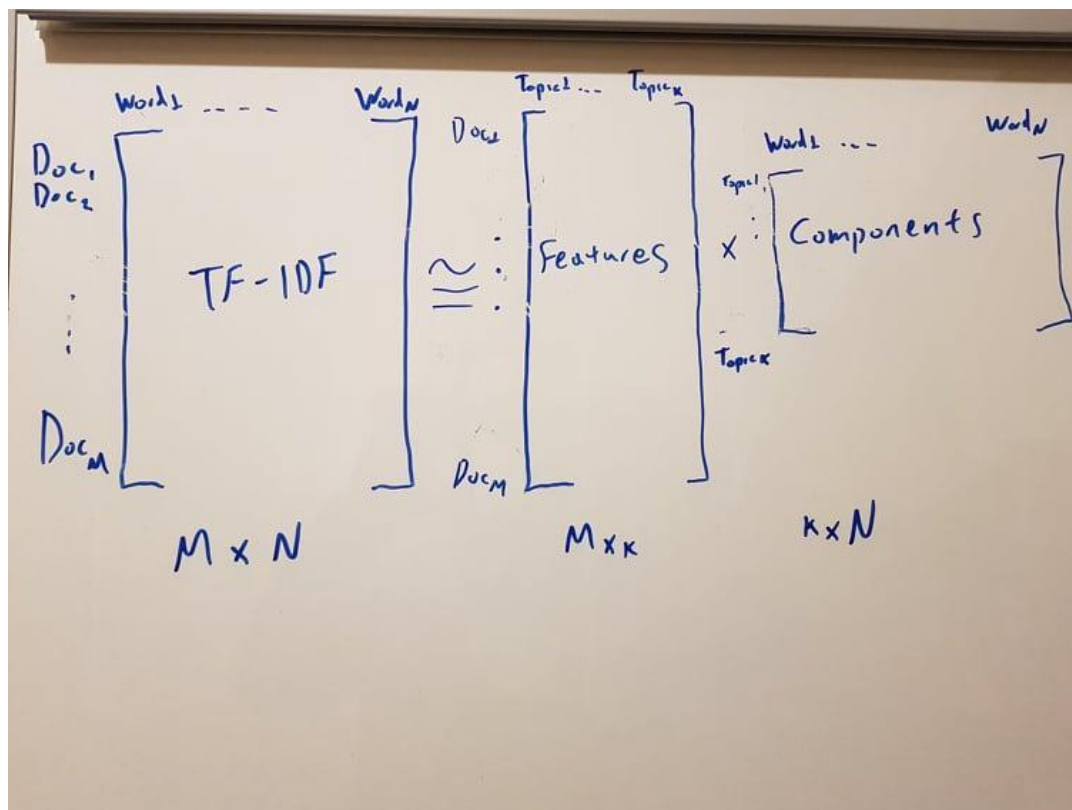
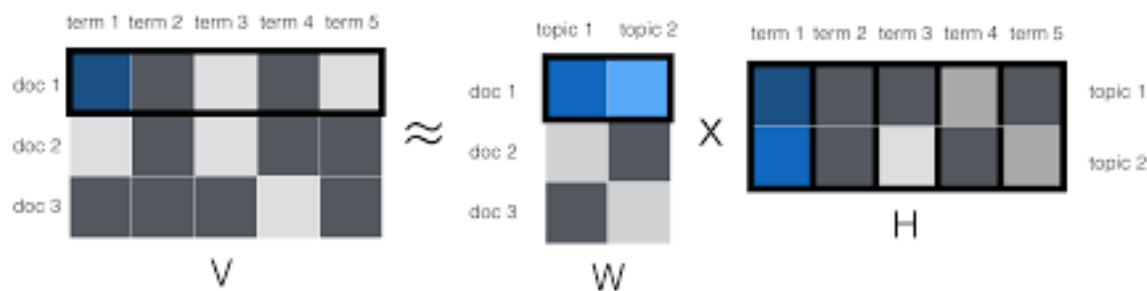
NMF is a technique for obtaining low rank representation of matrices with non-negative or positive elements.

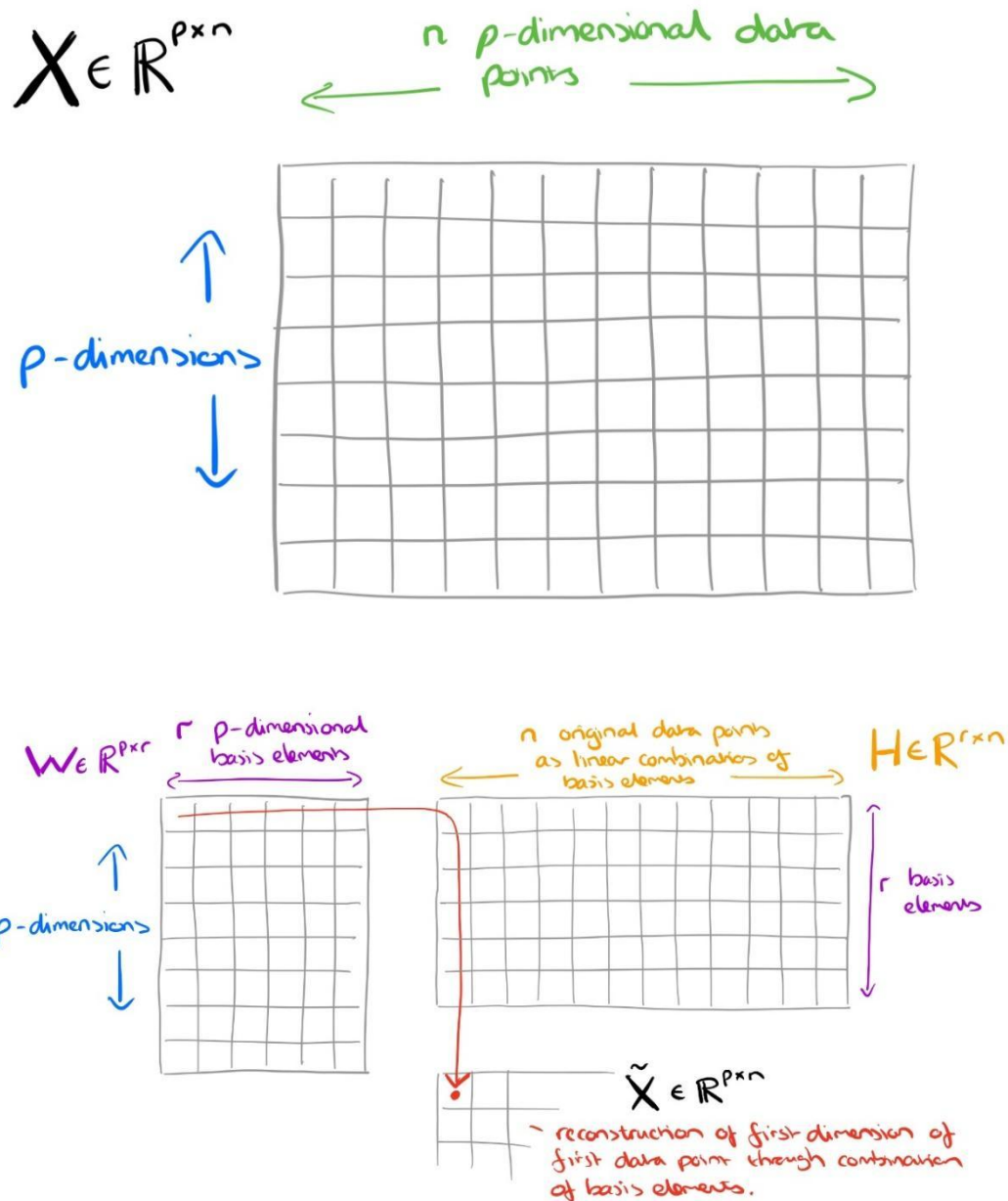
Similar to Principal component analysis (PCA) NMF takes advantage of the fact that the vectors are non-negative. By factoring them into the lower-dimensional form NMF forces the coefficients to also be non-negative.

How it works

Non-negative matrix factorization (NMF) relies heavily on linear algebra.

NMF approximates a matrix X with a low-rank matrix approximation (we want to reduce the P original dimensions to T) such that $X \approx WH$ where $X \in \mathbb{R}^{p \times n}$, $W \in \mathbb{R}^{p \times r}$ and $H \in \mathbb{R}^{r \times n}$ i.e. Given a data matrix A of p rows and n columns with each and every element $a_{ij} \geq 0$, NMF seeks matrices W and H of size p rows r columns and r rows n columns respectively such that $A \approx WH$ and every element of matrices W and H is either zero or positive.





A (Document-word matrix) - the input that contains which words appear in which documents. The input matrix A is the term counts or tf-idf matrix of size $p \times n$ where p is the number of documents and n is the number of terms. In other words, each element in the row represents countvectorizer or tfidf score of the term in the column for the respective document in the row.

W (Dictionary or Basis vectors) - the first decomposition output matrix W is the feature matrix of size $p \times r$, where p is the number of documents and r is the number of topic specified. In

other words, each element in the row represents the rank of a term of the document in the row for the respective topic in the column.

H (Expansion or Coefficient matrix) - the second decomposition output matrix H is the coefficient matrix of size $r * n$ where r is the number of topics specified and n is the number of terms. In other words, each element in the row represents the weight of the topic in the row for the respective term in the column.

The underlying idea of NMF is that a given data matrix A can be expressed in terms of summation of k basis vectors (columns of W) multiplied by the corresponding coefficients (columns of H). The matrices W and H are determined by minimizing the Frobenius norm i.e. a way of measuring how good the approximation WH actually is, is the Frobenius norm (denoted by the F subscript you may have noticed). The Frobenius norm is

$$||X - WH||_F^2 = \sum_{i,j} (X - WH)_{ij}^2$$

An optimal approximation to the Frobenius norm can be computed through truncated Singular Value Decomposition (SVD).

We calculate W and H by optimizing over an objective function and keep updating both W and H iteratively until convergence. Like most machine learning algorithms, NMF operates by starting with a guess of values for W and H , and iteratively minimizing the loss function. Typically, it is implemented by updating one matrix (either W or H) for each iteration and continuing to minimize the error function $||V - WH|| = 0$ (where W and H values remain non-negative) until W and H are stable.