**What is NNMF?**
- A type of low-rank approximation of a given matrix $A \in \mathbb{R}^{m \times n}$ where $a_{ij} \geq 0 \ \forall i, j$.
- Factors must be nonnegative.
- Certain applications (e.g., text mining, chemometrics, etc.) require nonnegativity in all the factors involved.
- SVD, PCA cannot be used because they involve negative coefficients, negative entries in the factors (i.e., entries of $U, V$ etc.).

**NNMF Objective**

Given a nonnegative matrix $A \in \mathbb{R}^{m \times n}$ and $k < \min(m, n)$, find nonnegative matrices $W \in \mathbb{R}^{m \times k}$, $H \in \mathbb{R}^{k \times n}$ to minimize the objective function

$$J_{NNMF}(W, H) := \frac{1}{2} \| A - WH \|_F^2$$

The product $WH$ is called an (approximate) NNMF of $A$.

The choice of $k$ is critical in practice, but often $k \ll \min(m, n) \Rightarrow$ a compressed approx. of $A$. 
**Numerical Approaches for NNMF**

- Minimization of $J_{NNMF}$ is difficult:
  - Many local minima exist in $J_{NNMF}$ in both $W$ & $H$.
  - Lack of a unique solution

  Consider $D \in \mathbb{R}^{k \times k}$, nonnegative and nonsingular, and suppose $D^{-1}$ is also nonnegative (e.g., $D$ could be $\text{diag}(d_1, \ldots, d_k)$ with $d_j > 0$ for $1 \leq j \leq k$.)

  Then if $WH$ is an NNMF of $A$, so is $WD^{-1}H$.

- Many algorithms have been proposed. We'll discuss only one of them based on the so-called Alternating Least Squares (ALS).

**Algorithm (ALS-NNMF)**

```plaintext
rand(n,k)
returns mxk
random matrix
whose entries
are uniformly
distributed
on the unit
interval (0,1).
```

- Initialize $W$ by $W = \text{rand}(m,k)$.
- For $j = 1 : \text{max iter}$
  - Solve for $H$ in $W^TWH = W^TA$.
  - Set all negative entries of $H$ to 0.
  - Solve for $W$ in $HH^TW^T = HA^T$.
  - Set all negative entries of $W$ to 0.

Compare it with "randn": the standard normal distribution!
Notes: (1) Convergence is not guaranteed yet this algorithm usually works in practice.
(2) \(W^TWH = W^TA, HH^TW^T = HA^T\)
are just a bunch of normal egn’s,
e.g., \(W^TWv_i = W^TA_i, \ i = 1:n\).
\(HH^TW^T = HA^T\) comes from the following:

\[\| A - WH \|_F^2 \rightarrow \min.\]
\[\iff \| A^T - H^TW^T \|_F^2 \rightarrow \min.\]
\[\iff (H^TH^T)W^T = (H^TA)^T\]
\[\iff HH^TW^T = HA^T.\]

(3) Random initialization like the original algorithm may not be efficient. We can use the following algorithm to initialize the matrix \(W\):

- Compute the first \(k\) singular values and the corresponding vectors by
\([U,S,V] = svds(A,k)\);
- Then do the following:
\(W(:,1) = U(:,1)\);
for \(j = 2:k\)
\(C = U(:,j) \times V(:,j)\);
\(C = C \times (C \geq 0)\);
\([u,s,v] = svds(C,1)\);
\(W(:,j) = u\);
end
The reasoning behind this initialization is the following:

If $A$ is nonnegative, then its first singular vectors $U_1$ & $U_1$ are also nonnegative. So, it's good to use $W(:, 1) = U_1$ and $H(1, :) = U_1^T$

Unfortunately, $U_2$, $U_2$ contains negative entries due to the orthogonality $U_1 \perp U_2$, $U_1 \perp U_2$.

So, construct $C = U_2 U_2^T$, and set all the negative entries of $C$ to 0. Then this $C$ is nonnegative, so can compute the first singular vectors of this $C$, which are nonnegative and good approximations to $U_2$, $U_2$. Then set the first left singular vector as the 2nd column of $W$. We can repeat this procedure until we fill $W$. 

Good exercise!
Example: Problem 2 of HW #1

Consider the following set of terms (words) and documents (or rather book titles):

<table>
<thead>
<tr>
<th>Terms</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2: Equation (Equations)</td>
<td>D2: NIST Handbook of Mathematical Functions</td>
</tr>
<tr>
<td>T3: Function (Functions)</td>
<td>D3: Table of Integrals, Series, and Products</td>
</tr>
<tr>
<td>T4: Integral (Integrals)</td>
<td>D4: Linear Integral Equations</td>
</tr>
<tr>
<td>T5: Linear</td>
<td>D5: Proofs from THE BOOK</td>
</tr>
<tr>
<td>T6: Mathematics (Mathematical)</td>
<td>D6: The Book of Numbers</td>
</tr>
<tr>
<td>T7: Number (Numbers)</td>
<td>D7: Number Theory in Science and Communication</td>
</tr>
<tr>
<td>T8: Series</td>
<td>D8: Green’s Functions and Boundary Value Problems</td>
</tr>
<tr>
<td></td>
<td>D9: Discourse on Fourier Series</td>
</tr>
<tr>
<td></td>
<td>D10: Basic Linear Partial Differential Equations</td>
</tr>
<tr>
<td></td>
<td>D11: Mathematical Physics, An Advanced Course</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term-Document Matrix</th>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$D_4$</th>
<th>$D_5$</th>
<th>$D_6$</th>
<th>$D_7$</th>
<th>$D_8$</th>
<th>$D_9$</th>
<th>$D_{10}$</th>
<th>$D_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>T6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Let's compute the NNMF of $A$ with $k=3$, using MATLAB:

$$[W, H] = nnmf(A, 3);$$
The resulting matrices are:

\[ W = \begin{bmatrix}
1.4366 & 0.0016 & 0 \\
0 & 1.4181 & 0 \\
0.9536 & 0 & 0 \\
0 & 0.6530 & 0.8984 \\
0 & 1.4181 & 0 \\
1.2931 & 0 & 0.0023 \\
0.4829 & 0.0076 & 0 \\
0 & 0 & 1.3883
\end{bmatrix} \]

\[ H = \begin{bmatrix}
0.2681 & 0.7569 & 0 & 0 & 0.2922 & 0.3875 & 0.0954 & 0.1967 & 0 & 0 & 0.2681 \\
0 & 0 & 0.0370 & 0.7573 & 0.0001 & 0.0042 & 0.0041 & 0 & 0 & 0 & 0.6520 \\
0.0014 & 0.0004 & 0.8342 & 0.1669 & 0 & 0 & 0 & 0 & 0 & 0.5251 & 0 & 0.0014
\end{bmatrix} \]

Let's interpret the results!

\( W_3 \) has large entries corresponding to

T4 (Integral/Integrals) and
T8 (Series).

The responses of the documents to \( W_3 \)
is the 3rd row of \( H \).

You can see that \( D_3 \) and \( D_9 \) have
high responses, which are reasonable:

\( D_3 = \) Table of Integrals, Series, and Products
\( D_9 = \) Discourse on Fourier Series

Exercise: Do interpret \( W_1 \) and \( W_2 \)
yourself!!