Introduction to Vector Space Models - Worksheet

Part One

1. Is the vector \( \mathbf{x} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \) in the \( \text{span} \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \} \)?

2. Is the vector \( \mathbf{x} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \) in the \( \text{span} \{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \} \)?

3. Describe the span of one vector in \( \mathbb{R}^3 \).

4. Describe the span of two linearly \( \text{independent} \) vectors in \( \mathbb{R}^3 \).

5. Describe the span of two linearly \( \text{dependent} \) vectors in \( \mathbb{R}^3 \).

6. Compare the \( \text{span} \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \} \) to the \( \text{span} \{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \} \).

7. What is the dimension of the \( \text{span} \{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \} \)?

8. What is the definition of the \textbf{dimension} of a subspace?

9. How would you describe a hyperplane?
Part Two

1. What are the coordinates of the vector \( x = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \) in the basis \( \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \)? Draw a picture to make sure your answer lines up with intuition.

2. In the following picture what would be the signs (+/-) of the coordinates of the green point in the basis \( \{v_1, v_2\} \)? Pick another point at random and answer the same question for that point.
Part Three

1. Interpret the following Nonnegative Factor Output for a small collection of text documents, answering the following questions:

   a. What meaning (theme/topic) would you give to each of the three factors?

   b. What is the dominant factor (theme/topic) for each document?

   c. What is the loading of the word *baseball* on Factor 2?

   d. What is the coordinate/score of document 5 along Factor 3?

\[
\text{TermDocMatrix} \approx \begin{pmatrix}
     \text{Factor1} & \text{Factor2} & \text{Factor3} \\
     \text{“baseball”} & 1.9 & 0 & 0 \\
     \text{“pitcher”} & 2.6 & 0 & 0.1 \\
     \text{“mound”} & 1.1 & 0.0 & 0 \\
     \text{“player”} & 1.5 & 0.1 & 0 \\
     \text{“coach”} & 1.3 & 0.8 & 0.8 \\
     \text{“soccer”} & 0 & 2.2 & 0 \\
     \text{“world”} & 0.1 & 1.7 & 0.5 \\
     \text{“fifa”} & 0 & 2.3 & 0 \\
     \text{“cup”} & 0 & 1.6 & 0.1 \\
     \text{“canada”} & 0.2 & 1.9 & 0.5 \\
     \text{“womens”} & 0 & 1.8 & 0.7 \\
     \text{“USA”} & 0.1 & 2.0 & 2.3 \\
     \text{“olympics”} & 0 & 0.2 & 2.8 \\
     \text{“medal”} & 0 & 0.1 & 2.2 \\
     \text{“gold”} & 0 & 0 & 1.8 \\
     \text{“phelps”} & 0 & 0 & 1.6 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
    \text{doc1} & \text{doc2} & \text{doc3} & \text{doc4} & \text{doc5} \\
    3.2 & 2.7 & 0 & 0.2 & 0.1 \\
    0.1 & 0.1 & 2.5 & 2.1 & 0.3 \\
    0.2 & 0 & 0.2 & 0.1 & 2.9 \\
\end{pmatrix}
\]

List of Key Words.

- linear combination geometrically
- linear (in)dependence geometrically
- vector span
- subspace
- dimension of subspace
- hyperplane
- basis vectors
- coordinates in different bases
- (generic) factor analysis
- loadings
- scores/coordinates