Introduction to Vector Space Models - Worksheet

Part One

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

   No. \( \mathbf{x} \) is not a scalar multiple of \( \begin{pmatrix} 1 \\ 1 \end{pmatrix} \).

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

   Yes. Since \( \begin{pmatrix} 4 \\ 3 \end{pmatrix} \) can be written as a linear combination of these vectors, it is in the span.

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

   A line through the origin.

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

   Equivalent answers:
   - A 2-dimensional subspace.
   - A plane through origin.
   - A hyperplane through origin.

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

   A line through the origin.

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

   They are exactly the same space! (a line)

7. What is the dimension of a subspace?

   The minimum number of vectors it takes to span the space.

8. How would you describe a hyperplane?

   A "flat" surface/subspace which cuts the ambient space in half.
Part Two

1. What are the coordinates of the vector \( \mathbf{x} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \) in the basis \( \left\{ \begin{pmatrix} -1 \\ 1 \\ 1 \\ 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ -2 \\ -1 \end{pmatrix} \right\} \)? 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 \( \{ \mathbf{v}_1, \mathbf{v}_2 \} \)? 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?
   - doc 1-2: Baseball (Factor 1)
   - doc 3-4: Women's WC (Factor 2)
   - doc 5: USA Olympics (Factor 3)

c. What is the loading of the word baseball on Factor 2? (which means baseball is not relevant to that factor)

   d. What is the coordinate/score of document 5 along Factor 3? 2.9

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

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

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