Introduction to Vector Space Models - Worksheet

Part One

1. Is the vector $\mathbf{x} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ in the *span* $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$?

   no. 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* $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$?

   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* $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ to the *span* $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$

   they are exactly the same space! (a line)

7. What is the **dimension** of a subspace?

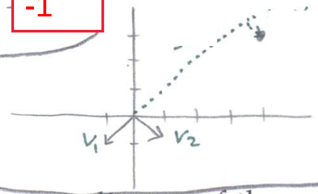
   the minimum number of vectors it takes to span the space.

8. How would you describe a hyperplane? a "flat" surface/subspace which cuts the ambient space in half.

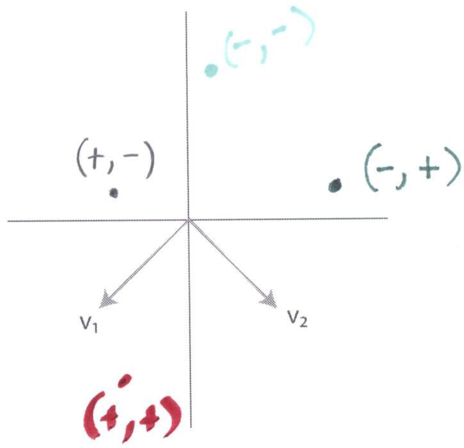
Part Two

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

$$\left(\begin{array}{cc|c} -1 & 1 & 4 \\ -1 & -1 & 3 \end{array}\right) \to \left(\begin{array}{cc|c} -1 & 1 & 4 \\ 0 & -2 & -1 \end{array}\right) \to \left(\begin{array}{cc|c} -1 & 0 & 3.5 \\ 0 & -2 & -1 \end{array}\right) \quad (-3.5, \tfrac{1}{2})$$



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 : Womens WC (Factor 2)
   doc 5 : USA Olympics (Factor 3)

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

   0 (which means baseball is not relevant to that factor)

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

   2.9

Baseball ⇑ Women's World Cup ⇑ USA Olympics Michael Phelps

$$TermDocMatrix \approx \begin{array}{c} \\ \text{"baseball"} \\ \text{"pitcher"} \\ \text{"mound"} \\ \text{"player"} \\ \text{"coach"} \\ \text{"soccer"} \\ \text{"world"} \\ \text{"fifa"} \\ \text{"cup"} \\ \text{"canada"} \\ \text{"womens"} \\ \text{"USA"} \\ \text{"olympics"} \\ \text{"medal"} \\ \text{"gold"} \\ \text{"phelps"} \end{array} \begin{array}{c} \textit{Factor1} \\ \left(\begin{array}{c} 1.9 \\ 2.6 \\ 1.1 \\ 1.5 \\ 1.3 \\ 0 \\ 0.1 \\ 0 \\ 0 \\ 0.2 \\ 0 \\ 0.1 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right. \end{array} \begin{array}{c} \textit{Factor2} \\ \begin{array}{c} 0 \\ 0 \\ 0.0 \\ 0.1 \\ 0.8 \\ 2.2 \\ 1.7 \\ 2.3 \\ 1.6 \\ 1.9 \\ 1.8 \\ 2.0 \\ 0.2 \\ 0.1 \\ 0 \\ 0 \end{array} \end{array} \begin{array}{c} \textit{Factor3} \\ \left.\begin{array}{c} 0 \\ 0.1 \\ 0 \\ 0 \\ 0.8 \\ 0 \\ 0.5 \\ 0 \\ 0.1 \\ 0.5 \\ 0.7 \\ 2.3 \\ 2.8 \\ 2.2 \\ 1.8 \\ 1.6 \end{array}\right) \end{array}$$

$$\begin{array}{c} \textit{doc1} \quad \textit{doc2} \quad \textit{doc3} \quad \textit{doc4} \quad \textit{doc5} \\ \left(\begin{array}{ccccc} 3.2 & 2.7 & 0 & 0.2 & 0.1 \\ 0.1 & 0.1 & 2.5 & 2.1 & 0.3 \\ 0.2 & 0 & 0.2 & 0.1 & 2.9 \end{array}\right) \end{array}$$

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