CFCS Tutorial One

Miles Osborne

February 2, 2009

1. Given the following two vectors: $\mathbf{a} = [1, 2, 3]$ and $\mathbf{b} = [4, 1, 2]$. Compute:
   - The length (norm) of each vector.
   - The dot product of these two vectors.
   - The length of **a** - **b** and **b** - **a**.
2. Suppose all documents consist of just sentences using the following words: *the dog cat sat on mat barked meowed.* Represent the following documents:
   (a) *the dog barked*
   (b) *cat meowed*
   (c) *the cat sat on the mat*

   using vectors. You should also explain how your representation works.
3. Work out the lengths (norms) of your vectors.
4. For each vector, work-out the corresponding unit vector.
5. Now, work out the following distances between each vector:
   - Absolute distance.
   - Cosine angle.
6. Which documents are most similar to each other? Does this vary according to the distance metric?
7. If you changed your document representation, how would it affect our distances?
8. In reality, vector representations of documents can deal with millions of possible words. What would happen to your represention? Any ideas how you can make it more space efficient?