

# COMS W4705: Natural Language Processing (Fall 2018)

## Problem Set #4

Wenbo Gao - wg2313@columbia.edu

December 4, 2018

### Problem 1

Using the raw co-occurrence counts:

- Which word is the most similar to 'animal' using euclidean distance?  
'dog'.

$$\text{dis}(\text{dog}, \text{animal}) = \sqrt{(0-2)^2 + (4-3)^2 + (0-0)^2 + (4-3)^2 + (2-0)^2 + (2-3)^2} = \sqrt{11}$$

$$\text{dis}(\text{cat}, \text{animal}) = \sqrt{(4-2)^2 + (0-3)^2 + (0-0)^2 + (3-3)^2 + (3-0)^2 + (10-3)^2} = \sqrt{71}$$

$$\text{dis}(\text{computer}, \text{animal}) = \sqrt{(0-2)^2 + (0-3)^2 + (0-0)^2 + (5-3)^2 + (0-0)^2 + (5-3)^2} = \sqrt{21}$$

$$\text{dis}(\text{run}, \text{animal}) = \sqrt{(4-2)^2 + (3-3)^2 + (5-0)^2 + (0-3)^2 + (3-0)^2 + (4-3)^2} = 4\sqrt{3}$$

$$\text{dis}(\text{mouse}, \text{animal}) = \sqrt{(2-2)^2 + (10-3)^2 + (5-0)^2 + (4-3)^2 + (3-0)^2 + (0-3)^2} = \sqrt{93}$$

- Which word is the most similar to 'animal' using cosine similarity?

$$\text{dis}(\text{dog}, \text{animal}) = \frac{\vec{v}_{\text{dog}} \cdot \vec{v}_{\text{animal}}}{\|\vec{v}_{\text{dog}}\| \|\vec{v}_{\text{animal}}\|} = \frac{0 * 2 + 4 * 3 + 0 * 0 + 4 * 3 + 2 * 0 + 2 * 3}{\sqrt{0^2 + 4^2 + 0^2 + 4^2 + 2^2 + 2^2} \sqrt{2^2 + 3^2 + 0^2 + 3^2 + 0^2 + 3^2}}$$

## Problem 2

## Problem 3