CFCS1

Distances

Miles Osborne

School of Informatics
University of Edinburgh
miles@inf.ed.ac.uk

January 14, 2010

Motivation

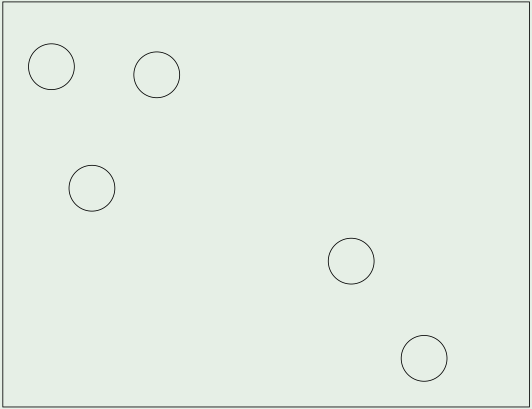
We often want to know how close vectors are to each other:

- Which Web page is similar to my query?
- Which person is like to me?
- Which neurons have a similar behaviour to each other?

We can capture these ideas using *distances* between vectors.

Motivation

Points in Space



Properties of distances

- There are many possible distance metrics.
- A correct distance metric has the following properties:
  - $d(x, y) \geq 0$ (Non-negativity)
  - $d(x, y) = 0$ if and only if $x = y$ (identity)
  - $d(x, y) = d(y, x)$ (symmetry)
  - $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)
- Any set such that these properties holds is called a *Metric Space*.
- Our cosine distance, over vectors, forms a metric space.
- The absolute value between vectors forms a metric space.
- (Later we will encounter the Kullback-Leibler Divergence, which is not symmetric)

Norms

- We first need an idea of how long a vector **a** is.
- The length of a vector is written: $\|\ \mathbf{a}\ \|$

$$\|\ \mathbf{a}\ \| = \sqrt{\sum_{i=1}^{n} a_i^2}$$

- Here the vector has components $a_1, a_2, \ldots a_n$.
- This follows from Pythagoras.

Norm Example

Suppose $\mathbf{a} = [-3\ 2\ 1]$
$\sqrt{(}9 + 4 + 1)$
$\|\ \mathbf{a}\ \| = \sqrt{14}$

Distance

- The distance between two points in space is another vector:

$$d(\mathbf{a}, \mathbf{b}) = || \mathbf{a} - \mathbf{b} ||$$

- Here we measure the length of the resulting vector.

Note:

- When the two vectors are identical, the distance is 0.
- The distance is always positive.

Examples

Distance Examples

$$d([1\ 3\ 2], [1\ 3\ 1]) = 1$$
$$d([1\ 3\ 2], [1\ 1\ 2]) = 2$$
$$d([1\ 3\ 2], [1\ 3\ 2]) = 0$$

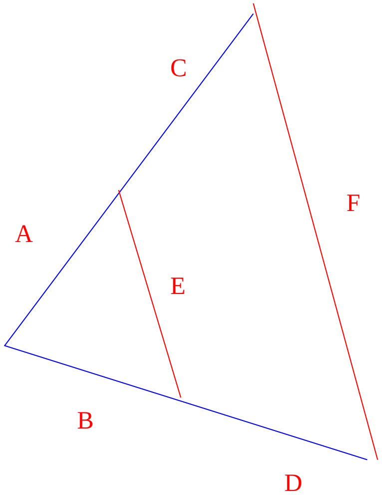
Cosine Distance

Suppose we multiply our vectors:

$$[1\ 3\ 2] * 3 = [3\ 9\ 6]$$
$$[1\ 3\ 1] * 3 = [3\ 9\ 3]$$
$$d([1\ 3\ 2], [1\ 3\ 1]) = 1$$
$$d([3\ 9\ 6], [3\ 9\ 3]) = 6.7$$

Our distance metric is sensitive to the absolute values of the components.

Cosine Distance



Cosine Distance

The angle between our two sets of vectors has however remained the same.

- The *cosine distance* measures the angle between two vectors.
- We can take the angle between two vectors and use that as a distance measure.

First we need to understand *dot products*.

Cosine Distance –Dot Products

The dot product of two **a** and **b** vectors is defined as:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

Dot Products Examples

$$[1\ 3\ 2] \cdot [1\ 3\ 1] = [1 * 1] + [3 * 3] + [2 * 1] = 12$$
$$[3\ 9\ 6] \cdot [3\ 9\ 3] = [3 * 3] + [9 * 9] + [6 * 3] = 100$$

Cosine Distance –Dot Products

Many vector ideas can be expressed using Dot Products:

- Lengths: $\| \mathbf{a} \| = \sqrt{(}\mathbf{a} \cdot \mathbf{a})$.
- Distances:$d(\mathbf{a}, \mathbf{b}) = \| \mathbf{a} - \mathbf{b} \|$.
- Angle (see next).

The properties of Dot Products determines the geometric properties of vectors in R^n.

Cosine Distance –Unit Vectors

We are still considering the absolute values of vectors.
The unit vector of vector of **a** is vector **b**:

$$\mathbf{b} = \frac{\mathbf{a}}{\| \mathbf{a} \|}$$
$$\| \mathbf{b} \| = \mathbf{1}$$

Unit vectors are normalised to have the same length.

Cosine Distance –Bringing it together

The cosine of two vectors **a** and **b**:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{|| \mathbf{a} || \cdot || \mathbf{b} ||}$$

Cosine Examples

$$[1\ 3\ 2]\quad [1\ 3\ 1]\,, \theta = 0.97$$
$$[3\ 9\ 6]\quad [3\ 9\ 3]\,, \theta = 0.97$$

- We are now taking the dot products of the corresponding unit vectors.
- As expected, the angles (distances) are the same.

Summary

- Distances between vectors are extremely useful.
- We looked at two distances:
  - Absolute distance.
  - Cosine distance.
- We also introduced the ideas of:
  - Norm (length) of a vector.
  - A dot product.
  - A metric space.