# Cosine Similarity in Terms of Euclidean Distance

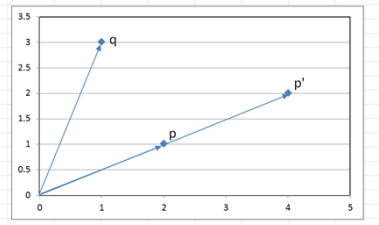
**Marzena Kryszkiewicz**
**Institute of Computer Science**
**Warsaw University of Technology**

1

## Intuition behind the Cosine Similarity

*The cosine similarity between vectors u and v* is denoted by *cosSim(u, v)* and is defined as the cosine of the angle between them.

*beautiful*




*flower*

2

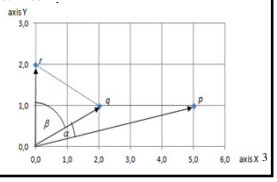
## Basic Notions: Cosine Similarity of Vectors

*The cosine similarity between vectors u and v* is denoted by *cosSim(u, v)* and is defined as the cosine of the angle between them; that is,

$$cosSim(u,v) = \frac{u \cdot v}{|u||v|}$$

- $u \cdot v$ is a standard vector dot product of $u$ and $v$ and equals $\sum_{i=1..n} u_i v_i$ ;
- $|u|$ is the length of vector u and equals $\sqrt{u \cdot u}$ .

**Example.**
$p$ is more cosine similar to $q$ than $r$,
$cosSim(p, q) = \cos\alpha$ is greater than
$cosSim(r, q) = \cos\beta$.



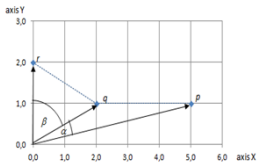

3

## The Cosine Similarity, the Triangle Inequality, the Normalized Form of a Vector

•The cosine similarity does not preserve the triangle inequality!

- $cosSim(u, v) = cosSim(NF(u), NF(v))$ where $NF(u) = \frac{u}{|u|}$

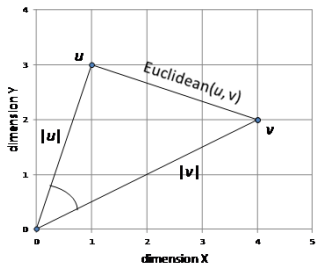
•The length of the normalized form $NF(u)$ of vector $u$ equals 1.




4

## Cosine Similarity and Euclidean Distance…

$$Euclidean^2(u, v) = |u|^2 + |v|^2 - 2|u||v|cosSim(u, v)$$




5

## Cosine Similarity and Euclidean Distance…

$$Euclidean^2(u, v) = |u|^2 + |v|^2 - 2|u||v|cosSim(u, v)$$

$$\Downarrow$$

$$cosSim(u, v) = \frac{|u|^2 + |v|^2 - Euclidean^2(u,v)}{2|u||v|}$$

$$\Downarrow$$

$$cosSim(u, v) = cosSim(NF(u), NF(v)) = \frac{2 - Euclidean^2(NF(u), NF(v))}{2}.$$

6

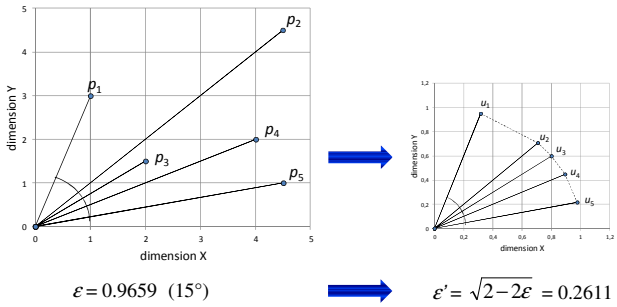
## Cosine Similarity and Euclidean Distance

$$cosSim(u, v) = cosSim(NF(u), NF(v)) = \frac{2 - Euclidean^2(NF(u), NF(v))}{2}.$$

$$cosSim(u, v) \geq \varepsilon \text{ if and only if } Euclidean(NF(u), NF(v)) \leq \varepsilon' = \sqrt{2-2\varepsilon}$$

7

## Example




$\varepsilon = 0.9659$ (15°)

$\varepsilon' = \sqrt{2-2\varepsilon} = 0.2611$

8

## References


- Marzena Kryszkiewicz: Determining Cosine Similarity Neighborhoods by Means of the Euclidean Distance. In: Rough Sets and Intelligent Systems (2013): 323-345
- Marzena Kryszkiewicz: The cosine similarity in terms of the Euclidean distance. Encyclopedia of Business Analytics and Optimization, 2498-2508


9