

## 2.9 几种常用的距离

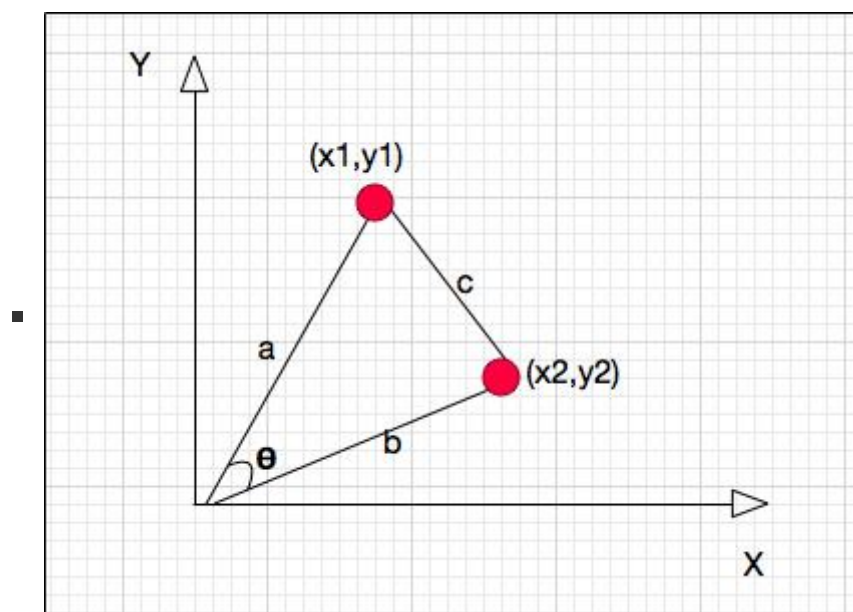
### 5. 余弦相似度

- 0. 基本解释:

- 公式:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

- 几何解释:



- 1. 解决问题: 通过测量两个向量的夹角的余弦值来度量他们之间的相似性
- 2. 取值: 取值在[-1,1]之间
- 3. 说明: 两个向量的夹角的余弦值确定两个向量是否大致指向相同的方向
- 4. 扩展: 上下界对任何维度都适用, 尤其是多维, 用处更广
- 5. 常用领域:
  - 文本挖掘中的文件比较
  - 数据挖掘中
- 6. python代码

#### # 2.9.5 夹角余弦

```
from numpy import *  
vector1 = mat([1,2,3])  
vector2 = mat([4,5,6]).T  
print(dot(vector1,vector2)/(linalg.norm(vector1)*linalg.norm(vector2)))
```

# 输出结果: [[0.97463185]]