

## 9.距离与相似度

在做分类时常常需要估算不同样本之间的相似性度量(Similarity Measurement)，这时通常采用的方法就是计算样本间的“距离”(Distance)。采用什么样的方法计算距离是很讲究，甚至关系到分类的正确与否。

### 1. 欧氏距离

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

### 2. 曼哈顿距离

$$d_{12} = |x_1 - x_2| + |y_1 - y_2|$$

### 3. 切比雪夫距离

$$d_{12} = \max(|x_1 - x_2|, |y_1 - y_2|)$$

### 4. 闵可夫斯基距离

$$d_{12} = \sqrt[p]{\sum_{k=1}^n |x_{1k} - x_{2k}|^p}$$

### 1. 夹角余弦

$$\cos \theta = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \sqrt{x_2^2 + y_2^2}}$$

### 2. 相关系数 & 相关距离

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(x)}\sqrt{D(Y)}}$$