# Machine Learning

## K-Nearest Neighbors

### 1. Distance metrics

**Euclidean distance**:

$$d(x, y) = \sqrt{\sum_{i=1}^{k} |x_i - y_i|^2}$$

**Cosine similarity**:

$$\cos(x, y) = \frac{x \cdot y}{||x|| \cdot ||y||}$$

**Jaccard distance**:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

**Manhattan distance**:

$$d(x, y) = \sum_{i=1}^{k} |x_i - y_i|$$

### 2. Model

**input**:

$$(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})$$

**for $\hat{x}$, find its k nearest neighbors**:

$$\hat{x}^{(1)}, \hat{x}^{(2)}, \cdots, \hat{x}^{(k)}$$

**find the majority class among these items**:

$$\hat{y} = \text{majority}\{\hat{y}^{(1)}, \hat{y}^{(2)}, \cdots, \hat{y}^{(k)}\}$$