定义：若样本点未知，则通过找与它距离最近的K个邻居样本点，通过 投票法/均值法 获取 样本的类别/值，完成对 样本的 分类/预测.

K近邻搜索作用 { 分类 / 预测

关键点：K 近邻搜索
- 如何确定K ①
- 如何定义距离从而找出近邻. ②

① 如何确定合适的K？

重要性：K过小 ⇒ 过拟合
K过大 ⇒ 欠拟合

Method: A：将K设置得大一些，但设置近邻投票权重（与距离成反比）

B：m折交叉验证，将K取不同值. 于每种值下进行m折交叉验证得到值，最后选出平均误差最小的K的取值

△ 此交叉验证方法在①决策树&&随机森林②岭回归与LASSO回归中均有使用.

② 如何确定距离？

Method: A: 欧式距离

$A: (x_1, x_2 \cdots x_n)$
$B: (y_1, y_2 \cdots y_n)$

$$d_{A,B} = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2 \cdots +(x_n-y_n)^2}$$

B: 曼哈顿距离.

$$d_{A,B} = |y_1 - x_1| + |y_2 - x_2| \cdots + |y_n - x_n|$$

C: 余弦相似度.

$$Similarity_{A,B} = \cos\theta = \frac{\vec{A} \cdot \vec{B}}{||\vec{A}|| \, ||\vec{B}||}$$

$$= \frac{x_1 y_1 + x_2 y_2 \cdots + x_n y_n}{\sqrt{x_1^2 + x_2^2 \cdots + x_n^2} \cdot \sqrt{y_1^2 + \cdots y_n^2}}$$

$\cos\theta$ ↑, θ ↓, 则相似度越大.

D: 闵可夫斯基距离公式.

$$d_{A,B} = \sqrt[p]{\sum_{i=1}^{n} (x_i - y_i)^p}$$

当 $p=1$ 时, 为曼哈顿距离;

$p=2$ 时, 为欧式距离;