1. **在k-means中，我们是用欧氏距离来计算最近的邻居之间的距离。为什么不用曼哈顿距离？**

欧氏距离虽然很有用，但也有明显的缺点。它将样本的不同属性（即各指标或各变量量纲）之间的差别等同看待，这一点有时不能满足实际要求。例如，在教育研究中，经常遇到对人的分析和判别，个体的不同属性对于区分个体有着不同的重要性。因此，欧氏距离适用于向量各分量的度量标准统一的情况。

曼哈顿距离，我们可以定义曼哈顿距离的正式意义为L1-距离或城市区块距离，也就是在欧几里得空间的固定直角坐标系上两点所形成的线段对轴产生的投影的距离总和。例如在平面上，坐标（x1, y1）的点P1与坐标（x2, y2）的点P2的曼哈顿距离为：，要注意的是，曼哈顿距离依赖座标系统的转度，而非系统在坐标轴上的平移或映射。当坐标轴变动时，点间的距离就会不同。

曼哈顿距离只计算水平或垂直距离，有维度的限制。另一方面，欧氏距离可用于任何空间的距离计算问题。因为，数据点可以存在于任何空间，欧氏距离是更可行的选择。

1. **线性分类器为什么要加入正则化项？**

不是满轶会导致线性回归无穷多解问题，而加入正则项可以解决该问题。

在训练数据不够多时，或者过度训练模型时，常常会导致过拟合。正则化方法即为在此时向原始模型引入额外信息，以便防止过拟合和提高模型泛化性能的一类方法的统称，这就引入了正则化惩罚项。在实际中，最好的拟合模型（从最小化泛化误差的意义上来说）是一个适当正则化的模型。