



链滴

# 李航《统计学习方法》第三章 k 近邻 python 实现 (CIFAR10 数据集)

作者: [hailangjiang](#)

原文链接: <https://ld246.com/article/1585663688537>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)



## KNN简述

k近邻法简单、直观，顾名思义即是对输入新样本在训练集中寻找k个特征最相似的样本，再根据这k个样本的标签通过多数表决等方式来预测新样本的类别。因此，k近邻不具有显示的学习过程。k近邻每预测都需要在整个训练集查找k个最近样本，实现查找最简单的方法是线性扫描，书中还讲述了一种对高效的实现方法kd树能够使查找实现在 $O(\log n)$ 时间量级，但由于时间限制本文只采取最简单的线扫描。k值的选择、距离度量及分类决策规则是k近邻法的三要素。

代码链接：[https://github.com/923012373/lihang\\_statistic\\_learning](https://github.com/923012373/lihang_statistic_learning)

## 数据集介绍

CIFAR10数据集，由10个类的60000个 $32 \times 32$ 三通道图片组成，每个类有6000张图像，有50000个训练样本和10000个测试样本。

## 实验部分

取k值为1，距离度量为L<sub>1</sub>距离，训练集测试集各10000，最后测试正确率为21.7%，这比随机10%效果要好了。

```
10000it [07:07, 23.38it/s]测试正确率:21.760000 %
```