

作业

1. 尝试在 MNIST 数据集上建立一个分类器，使它在测试集上的精度超过 97%。提示：`KNeighborsClassifier` 非常适合这个任务。你只需要找出一个好的超参数值（试一下对权重和超参数 `n_neighbors` 进行网格搜索）。
2. 写一个函数可以是 MNIST 中的图像任意方向移动（上下左右）一个像素。然后，对训练集上的每张图片，复制四个移动后的副本（每个方向一个副本），把它们加到训练集中去。最后在扩展后的训练集上训练你最好的模型，并且在测试集上测量它的精度。你应该会观察到你的模型会有更好的表现。这种人工扩大训练集的方法叫做数据增强，或者训练集扩张。

3. Tackle the Titanic dataset

The goal is to predict whether or not a passenger survived based on attributes such as their age, sex, passenger class, where they embarked and so on.

First, login to Kaggle and go to the Titanic challenge to download train.csv and test.csv。