KNN，全称是K-nearest neighbors。

KNN可以说是最简单的机器学习算法，其思想非常的简单。

就像它的名字一样，要对一个新的数据点做出预测，算法会在训练集中寻找与这个数据点距离最近的其他数据点，然后将找到的数据点的标签赋值给这个最新的数据点。

K最近邻中K的含义为，我们可以选择考虑训练集中与新数据点最近的任意K个邻居（例如可以观察1个，3个或是5个等等），而不是只考虑最近的那一个。

举个例子（熟悉的同学能够看出来是scikit-learn中鸢尾花iris数据集的一部分）：

图表, 散点图

描述已自动生成

以上的的每一个点都是一个数据点，每一个数据点有两个特征，分别是’sepal length’花萼长度以及’petal width’花瓣宽度，而黄色和紫色分别代表两种不同的鸢尾花的品种。

假设我们在野外观察到了一朵未知品种的鸢尾花，我们假设这朵花只有可能是这两种品种其一，KNN如何帮助我们预测此鸢尾花的品种呢？

例如我们发现了一朵鸢尾花的花萼长度是5.1 cm，花瓣宽度是1.6cm，那么例如我们观察离其最近的一个邻居时，我们发现最近的样本点的鸢尾花的品种是黄色代表的品种。于是1-NN分类器会将此朵鸢尾花也同样预测为黄色代表的品种。

如果我们观察离新数据点最近的三个邻居呢？我们会发现离此样本最近的三个邻居中，有一个是黄色代表的品种，而另外两个是紫色所代表的品种，根据投票法（voting），3-NN分类器将此朵鸢尾花预测为紫色所代表的品种。

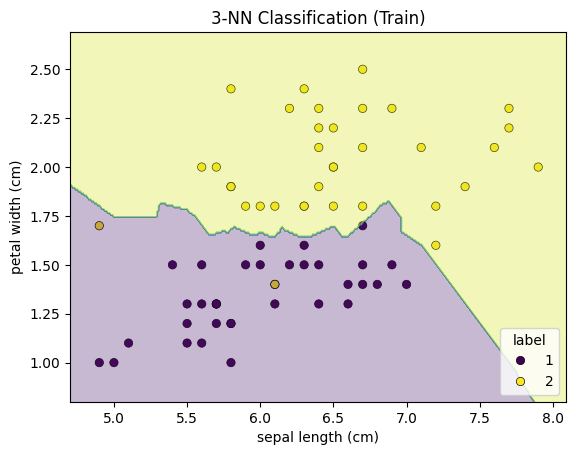
绘制决策边界

图表, 散点图

描述已自动生成

图表

描述已自动生成



图表

描述已自动生成

图表, 散点图

描述已自动生成

图表

描述已自动生成

图表, 散点图

描述已自动生成

图表

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图示

中度可信度描述已自动生成