**时序序列那点事儿**

分类篇（一）

**基于距离的分类：**

将样本根据其特征映射在空间中，根据在空间中的距离进行分类（K-NN）

Key Point:

如何根据特征将每个样本投影在空间中？

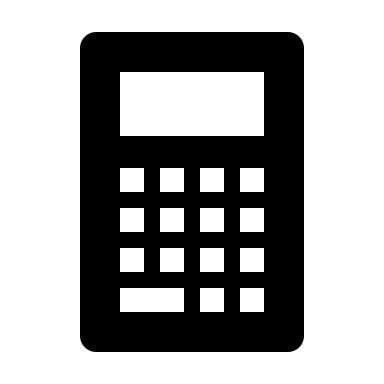
Or

如何比较两个样本之间的距离？

欧式距离，DTW

**基于特征的分类：**

基于样本的特征，直接找到对应的映射函数（可以是线性的也可以是非线性的），根据相对应的映射函数分类



F(T) (T = t1 , t2 …, tn)

Learning

Class

Key Point:

时间序列的本质（数据源的状态）应该是什么样子？

什么样的模型可以学习到时间序列的时序关系？

|  |  |
| --- | --- |
| Time (T) | value |
| t1 | v1 |
| t2 | v2 |
| … | … |
| tn | vn-1 |
| tn | vn |

然而，在某些情况下，对于有些样本，有多组时序数据，例如台风样本：

某个台风样本：

压强与时间的关系

温度与时间的关系

湿度与时间的关系

中心风速度与时间的关系

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time(T) | 压强 | 温度 | 湿度 | 中心风速 |
| t1 | p1 | temp1 | h1 | s1 |
| t2 | p2 | temp2 | h2 | s2 |
| … | … | … | … | … |
| tn-1 | pn-1 | tempn-1 | hn-1 | sn-1 |
| tn | pn | tempn | hn | sn |

更复杂的是

…

所以我们可以知道，一个样本在同一时间点上可能有多种类型的数据，每种类型的数据“分布”的形式不同，有些类型的数据就是一个数据点，有些类型的数据是就是个数据面，有些类型的数据是个数据立方体等等。

对于传统的分类问题往往数据源都是静态的，我们可以这样理解，在时序数据中

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time(T) | 压强 | 温度 | 湿度 | 中心风速 |
| t1 | p1 | temp1 | h1 | s1 |
| t2 | p2 | temp2 | h2 | s2 |
| … | … | … | … | … |
| tn-1 | pn-1 | tempn-1 | hn-1 | sn-1 |
| tn | pn | tempn | hn | sn |

在某一时间点t1下，由台风的各个属性来确定特征用来学习的，在很多机器学习的算法中，模型更看重的是同一样本中各个属性（特征）对样本的标签的影响，而对于时序数据则需要多考虑一步，那就是同一样本同一特征变化情况对样本标签的影响，在大部分处理时序数据的特征的算法中，都会将样本的特征的变化情况（时序依赖）抽取出来，这也是模型需要学习的。