

# 综合项目实践

## 基于机器学习的径流预报



# 选题背景

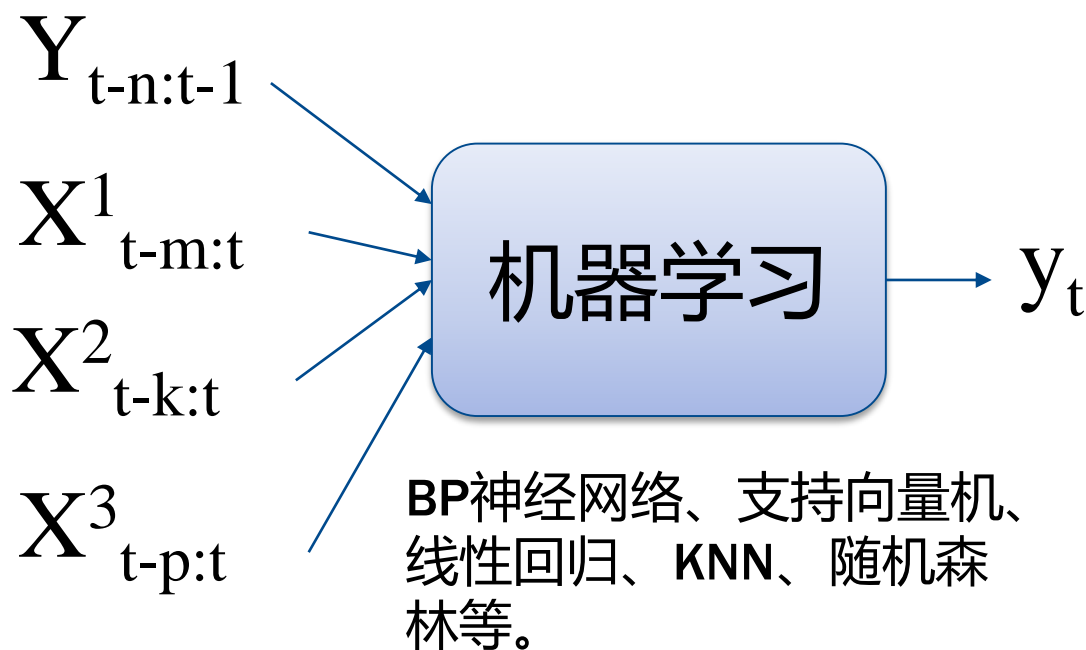
径流是指降水扣除蒸发等消耗后，沿地面、地下运动的水流。径流预报通常指对流域出口断面的流量作出预报。径流预报在防洪、抗旱、发电和灌溉中发挥着重要作用，也对水资源管理、生态环境保护具有重要意义。！



# 选题背景

径流预报的主要手段：传统的水文模型、机器学习方法。

随着人工智能技术的兴起和观测数据的积累，多种机器学习方法开始广泛应用于径流预报，并取得了较高的预报性能。



# 选题背景

基于机器学习的径流预报精度与机器学习算法本身、特征工程、建模空间尺度息息相关，此外通过集成学习技术能够进一步提高径流预报精度！

# 目标要求

**整体目标：**能够综合利用特征工程、机器学习和集成学习等技术实现径流预报。**具体目标**如下：

- 数据分析：了解数据集，进行数据统计分析和探索性数据分析。
- 特征工程：分别实现2种特征归一化方法和2种特征选择方法。
- 机器学习：选择2种机器学习算法实现给定单个流域的径流预报，并确定预报精度最优的特征工程方法+机器学习算法组合。
- 流域聚类：要求利用1种聚类算法将水文单元中的流域划分为不同的簇。
- 空间多尺度建模：针对给定水文单元内的多个流域，分别进行分区建模和全区建模，并确定预报精度最优的特征工程方法+机器学习算法组合。
- 集成学习：使用集成学习当中的Stacking和简单平均法将2种不同机器学习模型的预报结果进行集成。

# 任务流程

## 一、数据分析：

- 1、通过给定数据的网址或文档，理解所用数据集的内容。
- 2、对数据进行统计分析和探索性数据分析，并进行可视化，深入了解数据分布、异常及相关性等情况。

## 二、特征工程：

- 1、特征归一化：总结常用的特征归一化方法，并实现最小-最大归一化、Z-score标准化方法。
- 2、特征选择：总结常用的特征选择方法，并实现2种特征选择方法。

# 任务流程

## 三、机器学习回归：

- 1、调研适用于径流预报的机器学习回归算法。
- 2、实现2种机器学习算法（BP神经网络、支持向量机）。
- 3、利用Grid Search CV（网格搜索）调超参数。
- 4、基于给定的某个流域的数据，进行径流预报。
- 5、针对给定的单个流域数据，确定预报精度最优的特征工程方法+机器学习算法组合。

## 四、流域聚类：

- 1、使用聚类算法对给定水文单元内的多个流域进行聚类分析
- 2、确定最优的聚类数目，根据最优聚类数目将流域划归为不同的簇。

# 任务流程

## 五、空间多尺度建模：

1、分区建模：为给定区域的每个簇分别构建基于不同特征工程方法和不同机器学习的径流预报模型；利用Grid SearchCV调超参数；为每个簇确定预报精度最优的特征工程方法+机器学习算法组合，该组合称之为最优分区建模。

2、全区建模：为给定水文单元构建基于机器学习的径流预报模型。利用Grid SearchCV调超参数；为给定区域确定预报精度最优的特征工程方法+机器学习算法组合，该组合称之为最优全区建模。

3、分区建模与全区建模的精度对比：在同一套测试数据集上，对比最优分区建模和最优全区建模的径流预报精度。



# 任务流程

## 六、基于集成学习的径流预报：

- 1、实现基于Stacking的集成学习算法。
- 2、实现基于简单平均的集成学习算法。
- 3、基于给定的某个流域的数据，分别进行基于Stacking和简单平均的径流预报。
- 4、对比集成学习算法与单一机器学习算法的径流预报精度。

# 课程要求

- 自由组队，每队2人，指定一名小组长，分工合作完成题目；（小组长负责组织协调各位同学的工作，提交作业）
- 课程成绩：百分制
- 编程语言：Python
- 课程结束后，提交纸质版实践报告，电子版实践报告、汇报PPT、工程代码由小组长打包提交至网络教学平台【压缩包以“组长姓名+成员1姓名+成员2姓名”方式命名】

# 建议的开发环境与学习库

- 开源的Python发行版本Anaconda:

<https://zhuanlan.zhihu.com/p/32925500>

- Python集成开发环境PyCharm:

[https://blog.csdn.net/qq\\_44809707/article/details/122501118](https://blog.csdn.net/qq_44809707/article/details/122501118)

- 机器学习库Scikit-learn等

# 验收方式

- 分三次进行阶段验收，计入平时成绩，具体验收内容：
  - 第一次，机器学习径流预报验收
  - 第二次，多尺度空间建模验收。
  - 第三次，集成学习验收。
- 最后进行汇报验收，请预先做好汇报PPT，每组汇报时间10分钟以内，并回答问题；
  - 小组长、汇报人酌情加分