中国研究生创新实践系列大赛

中国光谷·“华为杯”第十九届中国研究生数学建模竞赛

题 目 草原放牧策略研究

摘 要：

草原作为世界上分布最广的重要的陆地植被类型之一，分布面积广泛。中国的草原面积为3.55亿公顷，是世界草原总面积的6%~8%，居世界第二。

问题1 从机理分析的角度，将提供的数据进行降噪处理，分析数据量纲，对其中多余无关数据进行处理，对土壤化学成分和物理成分数据进行分析，建立适当的数学模型。

问题2 通过统计学知识和原理，分析数据的均值、方差等，对应相应的卡方分布、t分布、F分布进行假设检验，不同统计方法进行预测不同条件下土壤给生物量带来的影响。

问题3 建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤化学性质影响的数学模型，接着对有机碳、无机碳、全N、土壤C/N比等值，这是对前面两问的一个模型扩展，继续采用前面的模型，在预测时需要结合放牧策略对化学性质的影响，需要对时间序列模型加上一个影响因子进行修正。

问题4 利用沙漠化程度指数预测模型和附件提供数据（包括自己收集的数据）确定不同放牧强度下监测点的沙漠化程度指数值，基于现有的模型和已有的数据代入求解。

问题5 可持续发展是经济规模增长未超过生物环境承载能力的发展。其理念是：“整个可持续发展理念，经济子系统的增长规模决不能超过生态系统永远可持续或支撑的容纳范围。”。

问题6 使用图案示例或动态演示方法分别预测2023年9月示范区的土地状况，使用Arcgis进行解答。

关键词：灰色关联分析、时间序列预测、目标规划、因子分析、Arcgis

目录

一、问题的背景与描述

中国草原主要分为温带草原、高寒草原和荒漠草原等类型。内蒙古锡林郭勒草原是温带草原中具有代表性和典型性的草原，是中国四大草原之一，位于内蒙古高原锡林河流，地理坐标介于东经 110° 50′ ~119° 58′ ,北纬 41° 30′ ~46° 45′ 之间，年均降水量340mm。内蒙古锡林郭勒草原不仅是国家重要的畜牧业生产基地，同时也是重要的绿色生态屏障，在减少沙尘暴和恶劣天气的发生方面发挥着作用，也是研究生态系统对人类干扰和全球气候变化响应机制的典型区域之一和国际地圈—生物圈计划（IGBP）陆地样带—中国东北陆地生态系统样带（NECT）的重要组成部分。

植物的生长满足自身的生长规律，同时受到周围环境的影响。例如，降水、温度、土壤湿度、土壤 PH、营养等都决定植物的生长情况。当牧羊（家畜日食量为 1.8kg，即标准羊单位，也包括羊羔。大牲畜折算系数 6.0，比如牛、马、骆驼，大牲畜幼崽折算系数 3.0 ）对植物进行采食时，一方面植物的地上生物量减少；另一方面，放牧对植物有刺激作用，改变了植物原有的生长速率，适当的放牧会刺激植物的超补偿生长，同样不合理的放牧也会降低植物的生长速率。

过度放牧，往往因牲畜密度过大，可能导致草原植被结构破坏，土壤裸露面积增大，促进了土壤表面的蒸发，土体内水分相对运动受到不利影响，破坏了土壤积盐与脱盐平衡，增加了盐分在土壤表面的积累，土壤盐碱化程度加重。最终造成草场退化、土地沙漠化。

土壤沙漠化又被称为沙质沙漠化，是荒漠化的一种主要表现类型。沙漠化是在干旱、半干旱和部分半湿润地区的沙物质基础和干旱大风动力条件下，由于自然因素或人为活动的影响，致使自然的生态系统平衡性遭到破坏，出现了以风沙活动为主要标志，并逐渐形成风蚀、风积地貌景观的环境退化过程，使原来没有沙漠景观的地区出现了类似沙漠景观的环境变化。所谓土壤板结化是指土壤打破了原有结构，表层的有机质遭到严重破坏而造成的，土壤板结原因很多，比如土壤贫瘠、容重过大、土壤质地太粘以及有机肥严重不足等。

问题 1. 从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤物理性质（主要是土壤湿度）和植被生物量影响的数学模型。

问题 2. 请根据附件 3 土壤湿度数据、附件 4 土壤蒸发数据以及附件 8 中降水等数据，建立模型对保持目前放牧策略不变情况下对 2022 年、2023 年不同深度土壤湿度进行预测，并完成下表。

问题 3.从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤化学性质影响的数学模型。并请结合附件 14 中数据预测锡林郭勒草原监测样地(12

个放牧小区)在不同放牧强度下 2022 年土壤同期有机碳、无机碳、全 N、土壤 C/N 比等值，并完成下表。

问题 4、利用沙漠化程度指数预测模型和附件提供数据（包括自己收集的数据）确定不同放牧强度下监测点的沙漠化程度指数值。并请尝试给出定量的土壤板结化定义，在建立合理的土壤板结化模型基础上结合问题 3，给出放牧策略模型，使得沙漠化程度指数与板结化程度最小。

问题 5、锡林郭勒草原近 10 的年降水量（包含降雪）通常在 300 mm ~1200 mm 之间，请在给定的降水量（300mm，600mm、900 mm 和 1200mm）情形下，在保持草原可持续发展情况下对实验草场内（附件 14、15）放牧羊的数量进行求解，找到最大阈值。（注：这里计算结果可以不是正整数）

问题 6、在保持附件 13 的示范牧户放牧策略不变和问题 4 中得到的放牧方案两种情况下，用图示或者动态演示方式分别预测示范区 2023 年 9 月土地状态（比如土壤肥力变化、土壤湿度、植被覆盖等）

二、问题的分析

问题 1 从机理分析的角度，建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤物理性质（主要是土壤湿度）和植被生物量影响的数学模型。首先我们要对提供的数据进行预处理，如异常值和多指标的去量纲处理。这是一种关联模型，找到放牧策略与土壤湿度和生物量的关系，可以采用灰色关联，因子分析模型，得到具体的各指标对湿度和生物量的权重。

问题 2 首先分析各指标的统计规律，首先要借助最基本的统计量，如趋势、中位数、方差等等，接着还可以使用配对样本 t 检验处理连续数据，卡方独立性检验处理离散数据。预测 2022 年、2023 年不同深度土壤湿度，采用时间序列 ARIMA 和支持向量机进行预测，并对多种模式进行组合，以提高预测精度，可以进行一个对比。

问题 3 建立不同放牧策略（放牧方式和放牧强度）对锡林郭勒草原土壤化学性质影响的数学模型，接着对有机碳、无机碳、全 N、土壤 C/N 比等值，这是对前面两问的一个模型扩展，继续采用前面的模型，在预测时需要结合放牧策略对化学性质的影响，需要对时间序列模型加上一个影响因子进行修正。

问题 4 利用沙漠化程度指数预测模型和附件提供数据（包括自己收集的数据）确定不同放牧强度下监测点的沙漠化程度指数值，基于现有的模型和已有的数据代入求解。土壤板结化与土壤有机物、土壤湿度和土壤的容重有关，建立各因素对之间的贡献权重，得到具体的板结化公式，其次根据第一二问的模型将物理化学性质反映到放牧方式，求出最优结果。

三、模型假设

1、数据真实可靠
2、土壤板结化因素只考虑土壤有机物、土壤湿度和土壤的容重

四、符号说明

符号	符号含义	符号	符号含义
$a1$	生物量	SL	增长率
$a2$	植被指数	GL	总贡献
$a3$	土壤蒸发量	P	绩效
$a4$	土地沙漠化指数	R	神经元数模
$a5$	降雨量	β	阈值学习速率
$a6$	土壤湿度	T	最大迭代次数
d	径流量	B	累加矩阵
K	分数指数	N	影响权重

五、模型的建立与求解

5.1 问题一

5.1.1 模型建立与求解

首先对附件数据进行预处理，进行等精度测量，独立得到 x_1， $x_2 \ldots$， x_n ，算出其算术平均值 $\bar{x}$及剩余误差 $v_i = x_i - \bar{x}$（ i=1,2，...,n）,并按贝塞尔公式算出标准偏差 δ，如果某个测量值 x_b的剩余误差 v_b（1<b<n）.满足下式：

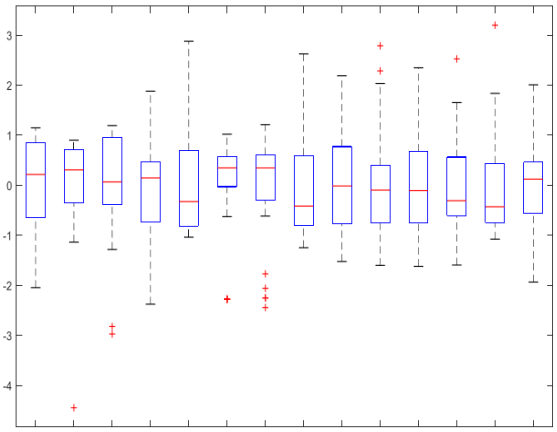
$$|v_b| = |x_b - \bar{x}| > 3\delta$$

我们认为 x_b是含有粗大误差值的坏值，应予以剔除。

对于一些不合常理的数据，我们假设这部分数据是因为手工写入差错或者抽样错误，并对这些数据进行适当处理，我们决定采取修改或者剔除的方法。我们假设给定指标的背景值范围限定在$x \pm 3\delta$区间内，因此我们认为在区间 $(x - 3\delta, x + 3\delta)$内的数据为正常范围的值，同时我们相对应的剔除异常数据。然后运用箱线图准确判断极端异常值的值，对这一部分数据，我们将其剔除并插入相应的值。

当数据当中存在异常值时， 尤其是存在着偏离较大的离群点时， 会对数据分析和建 模带来误差，因此，需要对异常值进行检测。常用的异常值检测方法包括 3σ 法则或 Z 分布方法， 而这一类方法是以正态分布为假设前提的。由于本文的一部分数值型特征的 分布并不符合正态分布， 即数据分布不均匀。故本文选择使用对数据分布没有要求的箱 线图，对数值型特性进行异常值检测。

使用箱线图对数据进行异常值检测的原理为：通过计算四分位数加减 1.5 倍四分位距，即是计算 Q1－1.5IQR 和 Q3+1.5IQR 的值，规定落在这一区间之外的数据为异常点。 在箱线图中， 可以看出变量数据的中位数、上四分位数、下四分位数、上下边缘和潜在 异常点。本文通过使用上四分位数代替数值大于 Q3+1.5IQR 的数据， 使用下四分位数 代替数值小于 Q1－1.5IQR 的数据，并绘制出了异常值处理前后的箱线图， 如图 所示。



按照灰色理论系统，把土壤湿度和植被生物量其余五种放牧方式和四种放牧强度作为变量，其它各个指标为比较数列 xi(x1、 x2，x3，x4，…)有公式计算出土壤湿度以及植被生物量与各指标之间的关联系数。

$$\xi i(k)=\frac{min_i\, min_k|y(k)-x(k)|+\rho max_i max_k|y(k)-x_i(k)|}{|y(k)-x(k)|+\rho max_i max_k|y(k)-x_i(k)|}$$

其中， $\xi i(k)$为 xi 对 y(k）在 k 点的系数：|y(k)-xi(k)|为第 k 点 y 与 xi 的绝对差; $\mathbf{min_i\, min_k|y(k)-x(k)|}$为 y 数列与 xi 数列在看点的二级最小差绝对值，$\mathbf{max_i max_k|y(k)-x_i(k)|}$为 y 数列与 xi 数列在看点的二级最大差绝对值，ρ 为灰色判别系数，取值在 0~1 之间，这里取值为 0.5，将各个指标的关联系数带入公式可得 xi 与 y(k）的关联度 ri

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi i(k)$$

5.1.2 描述性统计分析

为分析土壤湿度和植被生物量与放牧方式和强度，以及分析其变化统计规律， 本文绘制出两种因变量关于四个放牧强度的小提琴图， 如图所示。小提琴图刻画了数据的分布特征， 这种图用来显示数 据的分布和概率密度， 可以看成是箱线图和密度图的结合。小提琴图的中间部分反映箱 线图的信息， 图的两侧反映出密度图的信息。其中，小提琴图中间的黑色粗条用来显示 四分位数。黑色粗条中间的白点表示中位数，粗条的顶边和底边分别表示上四分位数和 下四分位数， 通过边的位置所对应的 y 轴的数值就可以看到四分位数的值。从小提琴图 的外形可以看到任意位置的数据密度，实际上就是旋转了 90 度的密度图。小提琴图越宽的地方表示数据密度越大，可以展示出数据的多个峰值。从图中可以看出，高强度的放牧 5-8 羊天/公顷会导致剧烈的土壤湿度降低，植被生物量也会遭受严重的破坏。对于轻度放牧，数据分布基本上只有一个明显的峰值，说明其湿度和植被生物量分布较为一致。对于中度放牧，数据分布呈多峰形态，分布呈现多级分化。

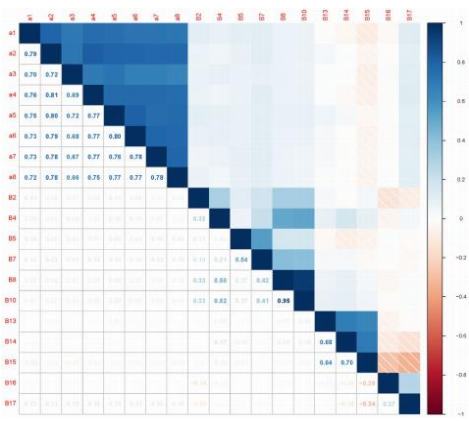
为了量化不同放牧强度在土壤湿度、植被生物量之间的具体差异，本文引入了 Wilcox 统计量和 Kruskal-Wallis 统计量[5]。其中， Wilcox 统计量用于比较两个总体分布之间的差异， Kruskal-Wallis 统计量用于比较 3 个总体分布之间的差异。在图中，展示出了统计量 的具体数值以及相应的 P 值。

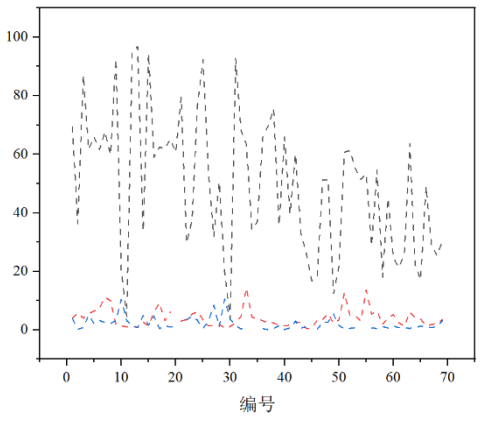
接下来， 本文进行了多变量的描述性统计分析， 绘制热力图， 如图所示。探究 了不同变量之间的相关性， 进而探究自变量之间的多重共线性。热力图， 又名相关系数 图， 根据热力图中不同方块颜色对应的相关系数的大小， 可以判断出变量之间相关性的 大小[6] 。两个变量之间相关系数的计算公式为：

$$\xi i(k) = \frac{min_i\, min_k|y(k) - x(k)| + \rho max_i max_k|y(k) - x_i(k)|}{|y(k) - x(k)| + \rho max_i max_k|y(k) - x_i(k)|}$$

公式中， ρ 表示相关系数， Cov 表示协方差， E 表示数学期望，即均值。

在热力图中，右侧的刻度展示了不同相关系数对应的颜色深浅。从图中可以看出，在进行特征工程时可以考虑剔除部分强相关的变量， 以免导致因多重共线 性造成的过拟合。





热力学图

5.1.3 典型性相关分析

假设放牧强度为U_i，土壤湿度为V_i，两组变量中选取若干有代表性的综合变量U_i，V_i，使得每个综合变量是原变量的线性组合，即

$$U_i = a_1^{(i)} X_1^{(1)} + a_2^{(i)} X_2^{(1)} + \cdots + a_P^{(i)} X_P^{(1)} \triangleq \mathbf{a}^{(i)} \mathbf{X}^{(1)}$$

$$V_i = b_1^{(i)} X_1^{(2)} + b_2^{(i)} X_2^{(2)} + \cdots + b_q^{(i)} X_q^{(2)} \triangleq \mathbf{b}^{(i)} \mathbf{X}^{(2)}$$

其中假设 $X^{(1)} = (X_1^{(1)}, X_2^{(1)}, \cdots, X_p^{(1)})$、$X^{(2)} = (X_1^{(2)}, X_2^{(2)}, \cdots, X_q^{(2)})$

当然，综合变量的组数是不确定的，为了有足够准确的数据，得保证第一组和第二组数据不相关，即

$\text{cov}(U_1, U_2) = \text{cov}(V_1, V_2) = 0$

利用 SPSS 可得下表：

表 1-7 典型相关性

	相关性	特征值	威尔克统计	F	分子自由度	分母自由度	显著性
1	.844	2.483	.194	17.970	9.000	168.078	0.000
2	.562	.463	.675	7.590	4.000	140.000	0.000
3	.110	.012	.988	.870	1.000	71.000	0.354

H0 for Wilks 检验是指当前行和后续行中的相关性均为零

假设显著性为 p，观察最后一列的 p 值，我们发现可将第一对和第二对进一步探究变量间的具体关系。查看“典型相关性表”，表中第一对、第二对典型变量的典型相关性不为 0，显著性 p =0.000<0.05，而剩余一组显著性大于 0，说明两组变量中共提取了一对典型变量，则说明在 99%的置信度水平下，结构变量和产品性能之间存在相关性，且第一对比第二对典型变量相关性极度显著，处于一个极端状态，因此选用第二对进行分析；第三对的 p 值为 0.354。说明第三对典型变量相关性不显著。

查看“集合 1”和“集合 2”标准化典型相关系数，集合 1 中第一对典型变量的标准化典型系数为（-1.092，0.021，-0.156），对应的，可以得到关系：

$$CV_{1\text{–}1} = -1.092 * y_1 + 0.021 * y_2 - 0.156 * y_3$$

类似地，观察集合 1 与集合 2 中的典型变量数对，可以得到：

$$CV_{1\text{–}2} = 1.642 * y_1 - 1.216 * y_2 + 1.265 * y_3$$

$$CV_{2\text{–}1} = 0.859 * Z_1 + 0.640 * Z_2 + 0.437 * Z_3$$

$$CV_{2\text{–}2} = 0.587 * Z_1 - 1.452 * Z_2 - 0.847 * Z_3$$

选用第二对后,将第一对和第三对数据删除后得下表:

集合 1 标准化典型相关系数

变量	1	2	3
放牧强度 mm	-1.092	1.642	-1.908

放牧方式（）	.021	-1.216	2.143
c	-.156	1.265	.135

集合 2 标准化典型相关系数

变量	1	2	3
土壤湿度	.187	.127	-.078
土壤蒸发量	.059	-.135	-.036
植被生物量	.005	-.010	-.016

5.2 问题二

5.2.1 时间序列分析法

我们采用时间序列分析法中的移动平均法，来预测收得率。

Step1： 设观测序列为时间序列，取移动平均的项数。一次移动的平均值计算公式为

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \ldots + y_{t-N+1})$$

$$= \frac{1}{N}(y_{t-1} + \ldots + y_{t-N+1}) + \frac{1}{N}(y_t - y_{t-N})$$

$$= M_{t-1}^{(1)} + \frac{1}{N}\left(M_t^{(1)} + \ldots + M_{t-N}^{(1)}\right)$$

Step2：二次移动平均值计算公式

$$M_t^{(2)} = \frac{1}{N}\left(M_t^{(1)} + \ldots + M_{t-N+1}^{(1)}\right) = M_{t-1}^{(2)} + \frac{1}{N}\left(M_t^{(1)} + \ldots + M_{t-N}^{(1)}\right)$$

Step3： 当预测目标的基本趋势是在某一水平上下波动时，我们可以用一次移动平均方法建立模型进行预测分析，即

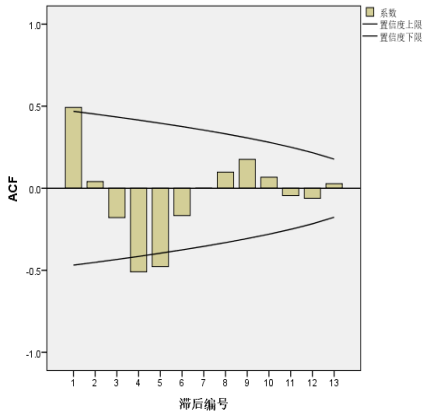
$$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \ldots + y_{t-N+1}) \quad t = N，N+1，\ldots，T$$

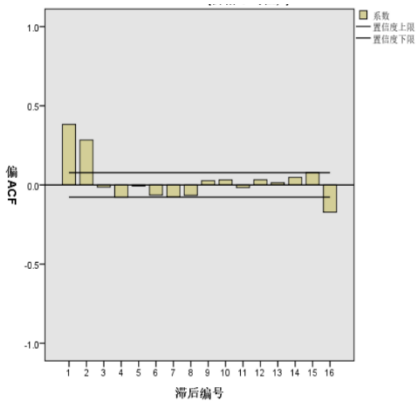
Step4:其预测标准误差为

$$S = \sqrt{\frac{\sum_{t=N+1}^{T}(\hat{y}_t - y_t)^2}{T-N}}$$

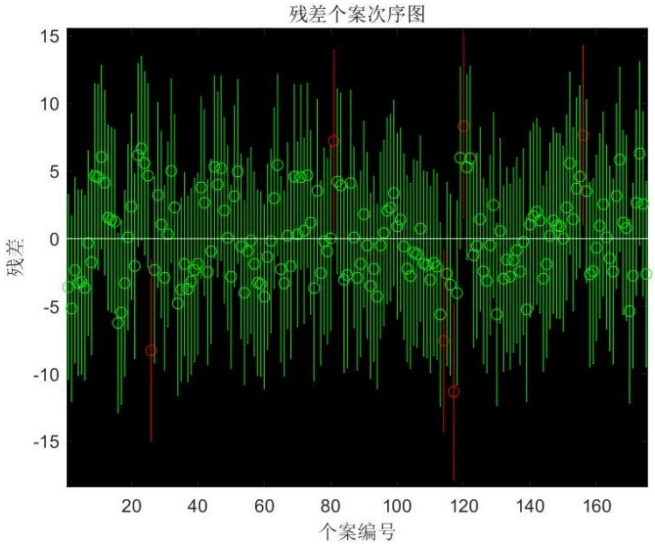
我们以最近 N 期序列值的平均值作为未来各期的预测结果。当历史序列的基本趋势变化不大且序列中随机变动成分较多时,N 的取值应该较大一些，否则 N 的取值应该小一些。在有确定的变动周期的资料中，移动平均的项数应选取周期长度。选择最佳 N 值的一个有效方法是,比较若干模型的预测误差，其中预测标准误差最小者为好。

借助 SPSS 软件进行求解，首先进行序列图分析，由于 P 值大于 0.05 我们对其做残差处理。





下表神经网络的预测结果，从后面的拟合优度和范数上不难看出预测结果良好。残差示意图及预测结果拟合如下：



BP 神经网络的搭建

BP 神经网络是一种按误差逆传播算法训练的多层前馈网络，类似于人工智能神经，它可以用来挖掘数据之间的未知关系，通过不断反向转播来不断调整网络的权值和阈值，目的在于使网络的误差平方和最小，BP 神经网络由一个输入层、一个或多个隐含层和一个输出层组成，每层含有若干个节，各层节点通过加权路径来与相邻层的节点进行链接。

BP 神经网络的工作方法主要分为三个过程。首先是信号的正向进行，信号从输入层加权计算传递到到隐含层，再通过隐含层的计算输出新的权重，如此循环，最后再到输出层；其次是误差的反向传播， BP 神经网络模型根据输出层的输出结果计算预测数据与实际数据的误差，再根据该误差调节从隐含层到输出层的权重和偏差以及输入层到隐含层的权重和偏差。最后计算准确率，进行模型测试，如此反复进行，直到输出结果与期望的输出结果均方误差达到一个合理的范围内则结束训练。

问题二可以看为湿度[X1]和降雨[X2]为输入,蒸发[Y1]，未来湿度[Y2]，不同深度湿度[Y3]为输出的复杂函数映射问题，采取隐藏层为五层，sigmod 函数为激活函数进行 BP 神经网络仿真。

Step1 训练集与测试集

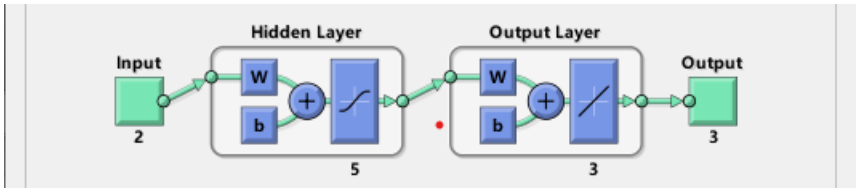
经过异常值的剔除和归一化处理之后，随机选择五个数据作为验证的样本，其余数据为训练集。

Step2 BP 神经网络的构建

Bp 神经网络包括输入层，五个隐藏层和输出层，神经元个数为 2 个，输出节点三个，从输出层到隐藏层以及隐藏层之间，采用 sigmoid 函数作为激活函数，从最后一个隐藏层到输出层之间采用 perelin 函数，得到网络输入数据与输出数据之间的表达式与下所示

$$y_k' = \sum_{j=1}^{r} v_j \cdot f[\sum_{i=1}^{m} w_{ij} \cdot p_i + \theta_j$$

其中(k=1,2,…N)，w_{ij}为链接权值，θ_j为阈值，y_k为期望输出值，y_k'为网络的实际输出值，神经网络结构的示意图如下图所示。



Step3 BP 神经网络的参数

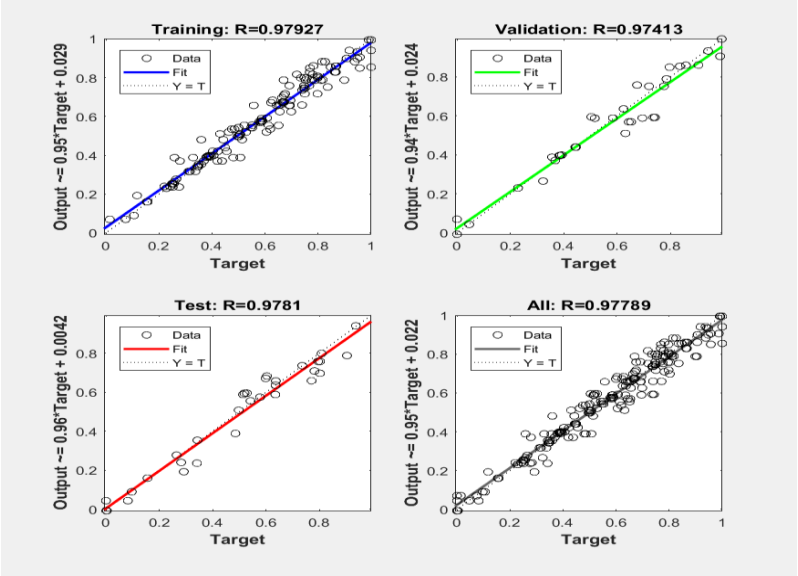
BP 神经网络进行预测主要包括的参数有：最大训练步数 net.tranParam.epochs，训练结果的间隔步数 net.trainParam.show，学习速率 net.trainParam.lr，训练目标误差 net.trainParam.goal，在本次模型中设置的参数如下所示：

最大训练步数	训练结果间隔步数	学习速率	训练目标误差	训练次数
1000	1	0.0000001	0.000001	1000

以 2022 年预测结果为例进行展示

年份	月份	10cm 湿度 (kg/m2)	40cm 湿度 (kg/m2)	100cm 湿度 (kg/m2)	200cm 湿度 (kg/m2)
2022	04	15.5	44.71	48.82	166.51
	05	14.09	43.84	47.06	166.72
	06	12.84	45.0	42.62	166.81
	07	10.55	40.65	62.65	167.03
	08	17.31	37.3	67.07	167.03
	09	17.13	32.99	55.32	165.92
	10	11.97	33.78	54.24	165.95
	11	11.76	33.85	46.04	166.09
	12	10.55	34.05	42.25	166.76
2023	**01**	**12.84**	**29.71**	**45.27**	**167.79**
	02	14.98	33.74	44.44	167.75
	03	13.72	35.79	44.52	167.72
	04	11.73	35.81	44.52	167.72
	05	14.09	42.71	55.02	168.32

	06	13.55	42.71	55.02	168.32
	07	12.85	42.7	55.02	168.32
	08	11.82	42.6	54.96	168.32
	09	12.73	40.66	54.08	168.32
	10	19.25	48.64	54.08	168.29
	11	20.34	54.64	62.91	168.22
	12	17.61	48.09	60.95	168.11



BP 神经网络预测模型的归回效果

从归回效果中可以看出，相关系数 R 值都高达 0.97 以上，说明 BP 神经网络的预测较为准确，并且范数为 0.7，精确度在 90%以上，可见预测性能良好，结果准确。

5.2.2 支持向量机模型

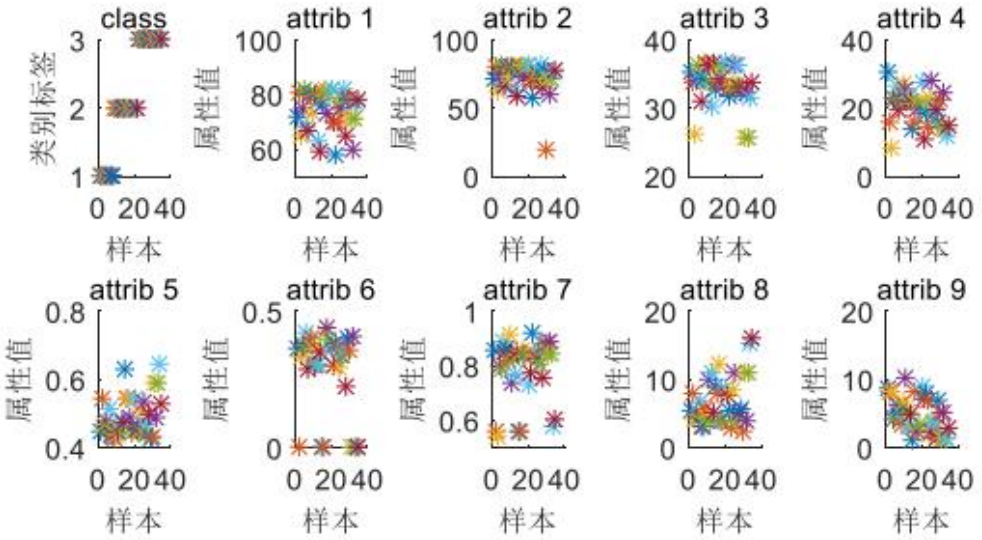
首先需要从原始数据里把训练集和测试集提取出来,然后进行定的预处理(必要的时候还需要进行特征提取),之后用训练集对 SVM 进行训练，最后用得到的模型来预测测试集的分类标签，流程图如下：

a)选定训练集和测试集

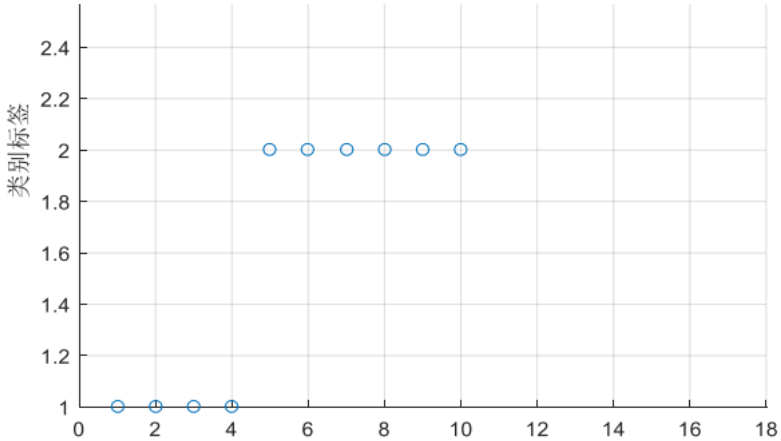
在这个 35 样本中,其中第 1-8 个样本属于第一类(类别标签为 1)，第 9-21 个属于第二类(类别标签为 2),第 22-35 个属于第三类(类别标签为 3)。现将每个类别分成两组，重新组合数据，一部分作为训练集(train_wine)，一部分作为测试集(test.wine)。

b)训练与预测

用训练集 train_wine 对 SVM 分类器进行训练，用得到的模型对测试集进行标签预测，最后得到分类的准确率。



可视化结果



因为只给了一个径流量，所以我们假设 Rin 与 Rout 相等，Gd 计算可参考题给文献（或者直接假设为 0）

下边界层的渗漏量。本研究将超出土壤最大有效贮水量的水分作为渗漏量来考虑，土壤最大有效贮水量的计算式为：

$$W_e = 0.1(W_f - W_c) \times \gamma \times n \qquad (3)$$

式中，W_e 为土壤有效贮水量(mm)；W_f 为一定土层

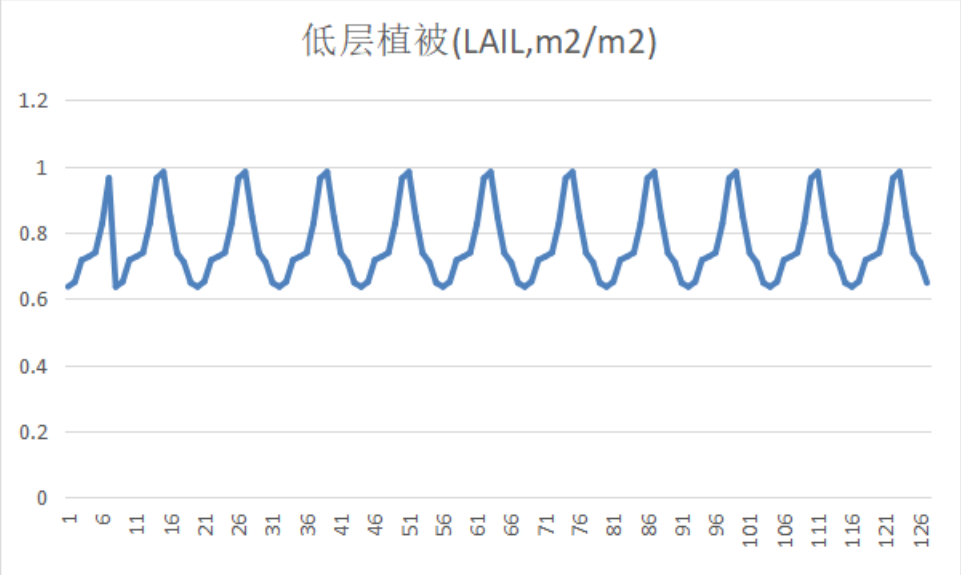
的田间持水量(%)；W_c 为一定土层的凋萎湿度(%)；γ 为土壤的平均容重(g/cm^3)；n 为土层厚度(cm)。

$$G_d = W_i - W_e \tag{4}$$

根据上面数据得到 ICstore(t)，

$$IC_{store}(t) = P(t) - E(t) - \triangle W + R_{in} - R_{out} - G_d$$

附件 10 叶面积指数 LAI：LAI(t) 已知，并且 LAI 为周期函数（下图），即未来月份 LAI 也已知。进一步推导 ICmax(t)已知



根据下式计算 k(t)

$$IC_{\text{store}}(t) = c_p \cdot IC_{\max}(t) \cdot [1 - \exp(-k(t) \cdot P(t)/IC_{\max}(t))] \tag{1}$$

k 为植被密度校正因子，与 LAI 有关；（对预测 k 有帮助，得到 k(t)后使用拟合方法得到 k 与 LAI 的关系，就能得到未来月份的 k 值）

降水量和蒸发量可通过 ARIMA 或者 SARIMA（季节性）时间序列预测

关于时间序列分析的p、q等参数，最优的p、q。可以通过一些准则（例如基于AICBIC准则的方法）确定。

求解步骤：

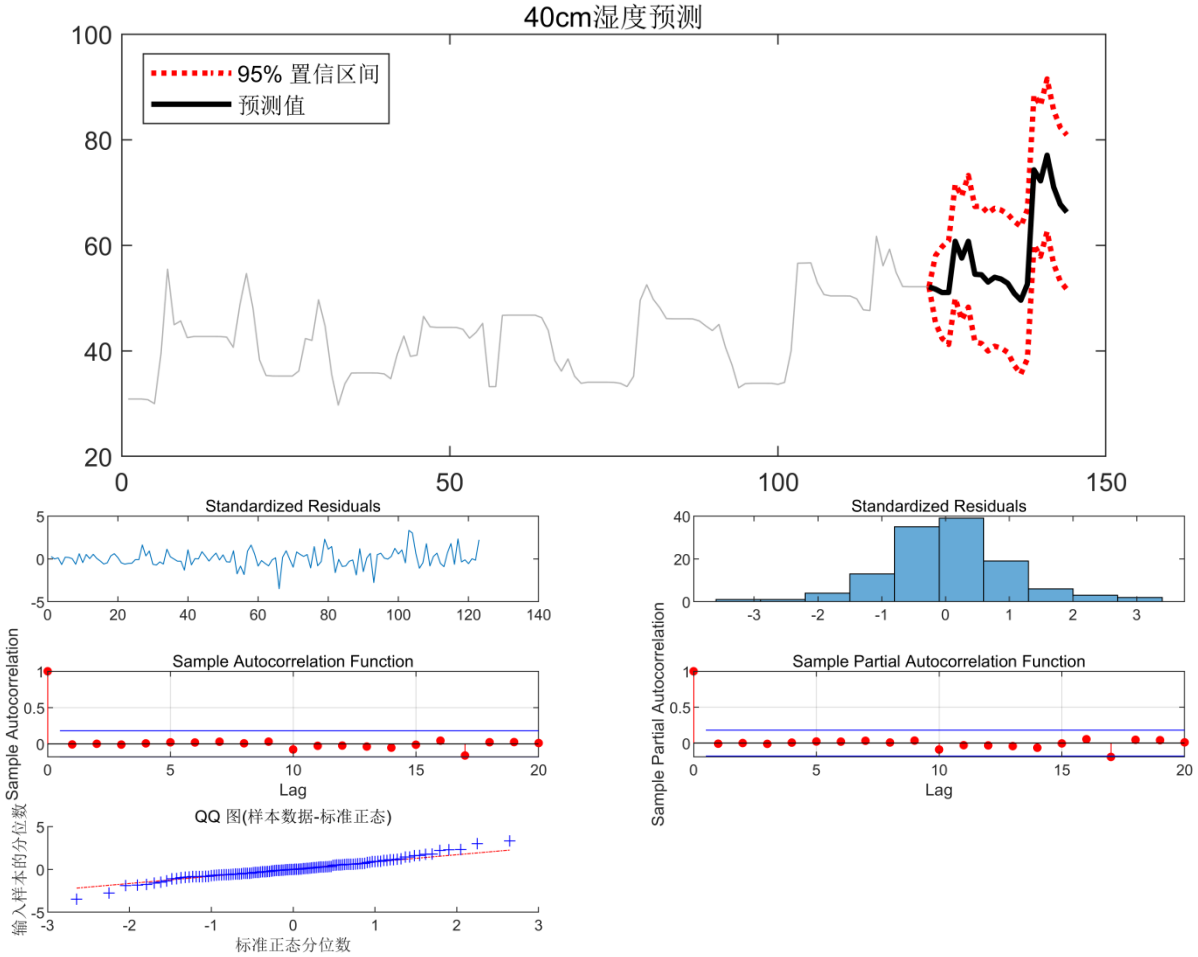
1. 通过 ARIMA 时间序列预测降水量、蒸发量、NDVI 等，需要注意季节性，使用季节性时间序列预测（SARIMA，见代码，SARIMA 的原理可以语文建模），使用时间序列预测前需要使用平稳性检验。检验方法有很多种，包括 ADF、KPSS、P-P 等。这里用 ADF 检验和 KPSS 检验。另外，Durbin-Watson 统计是计量经济学分析中最常用的自相关度量。该值接近 2，则可以认为序列不存在一阶相关性。
2. LAI 是周期函数，计算未来月份的 LAI，进一步计算未来月份的 ICmax
3. 计算未来月份的 ICstore(t)

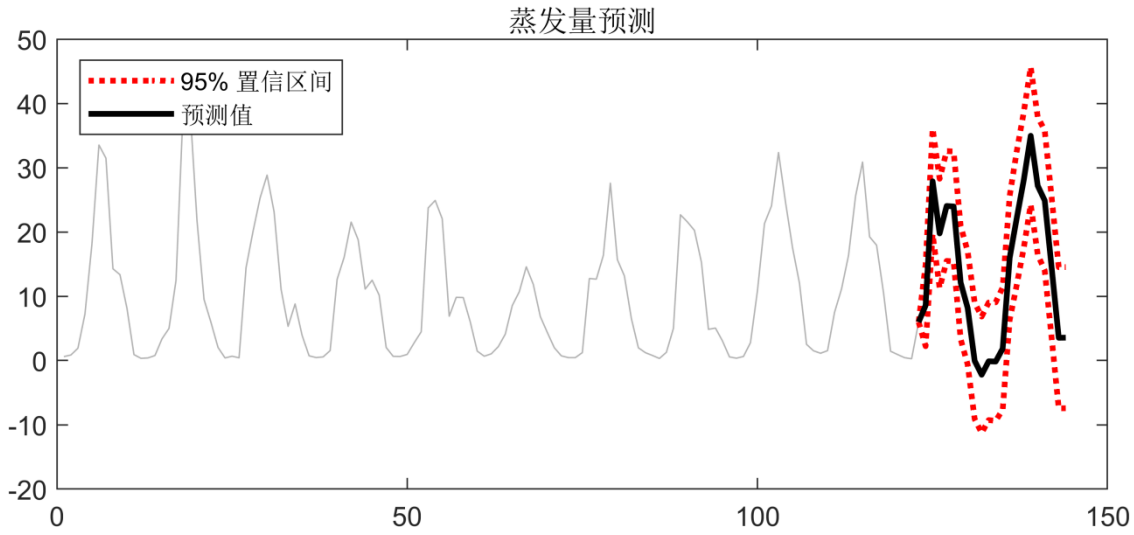
4. 根据未来月份的 ICstore(t)和预测的降水量和蒸发量，推导含水量变化
5. 通过含水量变化推导湿度变化

MATLAB 编程求解如下：

```
%% NDVI sarima 预测
AR_Order=15;
MA_Order=2;
SAR_Order=3;
SMA_Order=1;
S=12;
data=NDVI;
step=21;
savename='NDVI';
NDVI_forecast=sarima_forecast(AR_Order,MA_Order,SAR_Order,SMA_Order,S,data,st
ep,savename);
%% LAI 和 ICmax 预测
LAI_forecast=LAI(100:120);
ICmax_forecast=0.935+0.498*LAI_forecast-0.00575*LAI_forecast.^2;
```

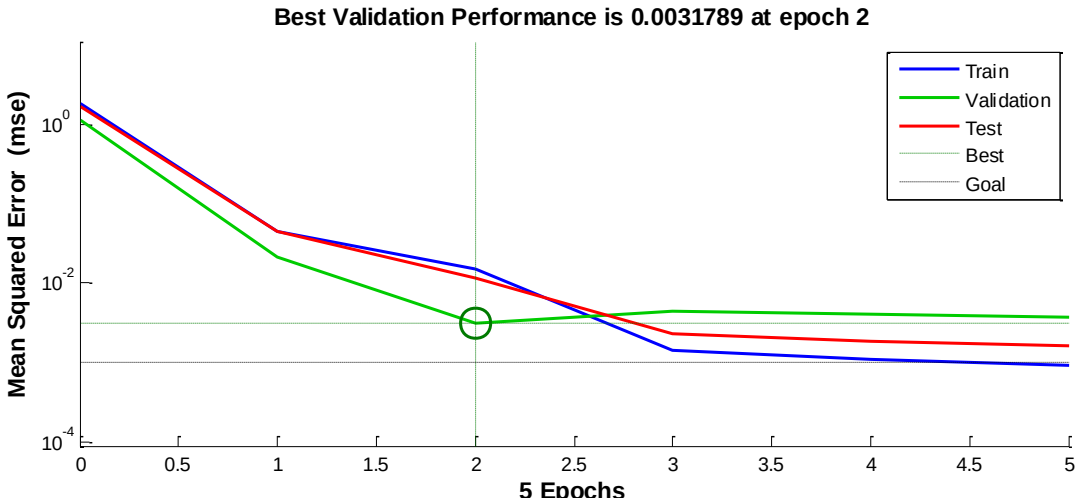
预测结果





模型结果

算法	类型数量	真实数量	正确率
支持向量机	69	60	86.7%



分类最终收敛，敏感性较好。

参考文献

[1] 肖郴杰. 基于深度学习的 P300 脑机接口分类算法研究[D].华南理工大学,2019.
[2]王磊. 基于运动想象的脑电信号分类与脑机接口技术研究[D].河北工业大学,2009.
[3]苏煜. 基于 SCF 范式的在线 P300 脑机接口研究[D].浙江大学,2010.
[4]黄小利. 脑电全脑信号及其在睡眠中的应用[D].西南大学,2019.
[5]程佳. 基于脑电信号的睡眠分期研究[D].北京理工大学,2015.
[6]李奇,卢朝华.基于卷积神经网络的 P300 电位检测及在脑机接口系统中的应用[J].吉林师范大学学报(自然科学版),2018,39(03):116-122.
[7]肖郴杰. 基于深度学习的 P300 脑机接口分类算法研究[D].华南理工大学,2019.
[8]单海军. 基于运动想象的脑机接口通道选择算法研究[D].浙江大学,2015.
[9]. Science - Applied Sciences; Research from Jilin University Provides New Data on Applied Sciences (Assessment of Landslide Susceptibility Combining Deep Learning with Semi-Supervised Learning in Jiaohe County, Jilin Province, China)[J]. Science Letter,2020.
[10]张锦涛. P300 脑机接口的在线半监督学习算法与系统研究[D].华南理工大学,2015.
[11]沈之芳. 基于 P300 的脑机接口及其在线半监督学习[D].华南理工大学,2014.
[12]刘佳.数学建模中的主成分分析法[J].科技视界,2014(15):223-224.
[13]张剑飞,王真,崔文升,刘明.一种基于 SVM 的不平衡数据分类方法研究[J].东北师大学报(自然科学版),2020,52(03):96-104.