**生态学复习要点**

第一章 绪论

1. 生态学主要采用的是“假说-验证”范式，面临的困境是什么？

2. 何为“数据科学”范式，与“假设-验证”范式主要区别是什么？

3. 什么是数据挖掘？与传统数据分析有什么区别？

第二章 R语言与数据挖掘

1. 为何R语言被用于数据挖掘主流工具之一？它具有哪些主要特点？

2. 机器学习建模与传统统计模型有哪些不同？（从算法和评估模型方面）

3. 什么是贝叶斯理论？为何可用于机器学习建模的？

4. 生态学上常用树模型，包括随机森林、提升回归树，可用于分类和回归。

1）有哪几个重要参数？建模过程中，采用何种方法优化这些参数？

2）分别给出评价分类、回归模型的具体指标是什么？

5. caret包提供了统一建模框架，包括分割数据，选择特征，评估模型等。

1）一般将数据集分割训练集与测试集，用到哪个函数？

2）优化模型参数用到验证集，该数据集包含在训练集或是测试集？

3）什么是选择特征，如何选择的，用到caret包中哪个函数？

4）通过什么函数指定优化参数方法？

第三章 共享数据与数据库

1. 生态学数据可分为生物属性数据与环境数据，主要数据来源是什么？

2. 遥感影像是重要共享数据源，举例说明生态学上可利用遥感影像获取哪些数据？

3. 生态学常用的数据库是postgresql或postgis，QGIS自带的是哪种？可存储非空间数据？

4. 什么是数据共享？主要共享平台或途径是什么？

第四章 探索性数据分析

1. 探索性数据分析主要包括哪些主要内容，其目的是什么？

2. 群落研究中，针对样方的分析（Q-mode）是基于距离、相关系数或协方差？根据此结果的聚类，有哪几种主要聚类方法？

3. 针对物种或栖息地环境变量分析（R-mode）是基于距离还是相关系数、协方差？根据此结果进行排序，排序的方法有哪些？

4. 排序可依据线性模型或是单峰模型，主要判断方法是R中哪个包或函数？限制性排序与非限制性排序主要区别是什么？

5. 在双序图中，点表示样方，矢量表示物种或环境，图解排序结果要看标度，当scaling=1，矢量长度和夹角分别表示的是什么？当scaling=2，矢量长度和夹角分别表示的是什么？

第五章 空间数据探索与建模

1. 空间数据有哪些类型？样方、研究区域、地形地貌分别以何种类型数据存储的？

2. 操作空间数据可以用QGIS和R，两者可以互用，在R中有哪些包在两者之间建立沟通的桥梁，举例说明。

3. 原始卫星影像资料要经过哪些处理步骤，才能用于生态学有关分析？哪些层级的遥感资料可直接使用？其主要用途是什么？

4. 利用空间数据，可绘制静态研究区域地图，也可建立预测模型，能否就全球变化导致物种分布改变，说明具体的建模步骤？

第六章 时间序列数据与建模

1. 时间序列数据纬度高，处理难度大，通常要降维，目前降维的主要方法有哪些？R中有哪些包可帮助实现重表达？

2. 不同于表格数据，时间序列数据通常只有测定时间和测定值，没有建模用的特征，需要自己构建特征，一般从时间戳或数值本身创建，基于数值创建特征的方法有哪些？

3. 基于数值特征可构建预测模型，其主要依据是什么？如何确定或判读这个依据是否存在？

第七章 生态网络数据与建模

1. 生态网络主要有one-mode和two-mode类型，主要区别是什么？可用矩阵形式表达，请举例说明。

2. 生态网络特性（properties）与建模特征（features）是否相同，请举例说明生态网络有哪些主要特征？

3. 建立网络链接模型，基于哪些顶点及其特征，基本原理是什么？建模步骤有哪些？