**Homework-05**

问题：

1. 从建模目标、建模算法和评估方面，阐述机器学习建模与传统统计模型有哪些不同？

2. 生态学上常用树模型，包括随机森林、提升回归树。建模步骤包括：数据预处理、拆分数据集、选择特征、算法和训练模型、模型评估等。caret为各种机器学习算法提供了统一模板，加载mtcars数据集，请根据问题填空。

1）对于doubs中的鱼群数据，按照样地，计算各样地鱼类Shannon多样性指数，并新增mpg列。

2）利用train()，训练随机森林（randomForest）模型

model\_rf <- train(mpg ~ ., data = training\_data, method = " \_\_")

3）通过trainControl()，向train()添加重采样10-fold cross-validation，以优化参数

fitControl <- trainControl(method = " ", number = 10, repeats = 5)

model\_rf <- train(mpg ~ ., data = training\_data, method = " ", trControl =fitControl)

3）在train()中，增加中心化和标准化等数据预处理，提高模型精度

model\_rf <- train(mpg ~ ., data = training\_data, method = " ",

preProcess = c('scale', 'center'),

trControl =fitControl)

4）rf有mtry和tree两个参数，可以通过expand.grid()设置调优，并在train()添加

grid <- expand.grid(.mtry=c(1:10))

model\_rf <- train(mpg ~ ., data = training\_data, method = " ",

preProcess = c('scale', 'center'),

trControl =fitControl,

\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_)

3. 什么是递归消除选择？在caret包中，为何选择随机森林等树模型时，没有特征选择这个过程？

要求：

回答问题时，将答案写为Word文档，包括填空内容，并命名为hwk-05，一周内上传代码至自己GitHub托管的homework中。