Homework 1
Econ 7720
Due Sep 11, 2023

The first task is to be familiar with Python (Jupyter Notebook) in a k-Nearest Neighbor algorithm application using the Abalone dataset. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope – a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. The question is to infer the age (ring) using easy measurements, e.g., length, diameter, height, whole weight, shucked weight, viscera weight, and shell weight. Please build a k-Nearest neighbor model to finish this task and have a report with the model, model performance through train-test split, and argument of choosing your number of neighbors.

Sample code to import the Abalone data

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
column_names = ["sex", "length", "diameter", "height", "whole weight",
                "shucked weight", "viscera weight", "shell weight", "rings"
                                                                    ]
abalone= pd.read_csv("abalone.data", names=column_names)
print("Number of samples: %d" % len(abalone))
abalone.head()
for label in "MFI":
    abalone[label] = abalone["sex"] == label
del abalone["sex"]
X = abalone.drop("rings", axis=1)
X = X.values
y = abalone["rings"]
y = y.values
```