Forest Coverage Prediction via Statistical Learning Models
STAT 6494 Data Science Project Proposal

Tairan Ye

March 8, 2018

1 Introduction

Natural resource managers responsible for developing ecosystem management strategies require basic descriptive information including inventory data for forested lands to support their decision-making processes. However, managers generally do not have this type of data for inholdings or neighboring lands that are outside their immediate jurisdiction. One method of obtaining this information is through the use of predictive models.

2 Data

The dataset is named "Forest Cover Type Dataset", from the "UCI Machine Learning Repository". It contains 12 predictors and 581,012 observations. The more detail information about the variables are summarized as following:

Name	Data Type	Measurement	Description
Elevation	quantitative	meters	Elevation in meters
Aspect	quantitative	azimuth	Aspect in degrees azimuth
Slope	quantitative	degrees	Slope in degrees
Horizontal Distance To Hydrology	quantitative	meters	Horz Dist to nearest surface water features
Vertical Distance To Hydrology	quantitative	meters	Vert Dist to nearest surface water features

Table 1: Summary of variables

Name	Data Type	Measurement	Description
Horizontal Distance To Roadways	quantitative	meters	Horz Dist to nearest roadway
Hillshade 9am	quantitative	0 to 255 index	Hillshade index at 9am, summer solstice
Hillshade Noon	quantitative	0 to 255 index	Hillshade index at noon, summer soltice
Hillshade 3pm	quantitative	0 to 255 index	Hillshade index at 3pm, summer solstice
Horizontal Distance To Fire Points	quantitative	meters	Horz Dist to nearest wildfire ignition points
Wilderness Area	qualitative	0 (absence) or 1 (presence)	Wilderness area designation
Soil Type	qualitative	0 (absence) or 1 (presence)	Soil Type designation
Cover Type	integer	1 to 7	Forest Cover Type designation

Table 2: Summary of variables

3 Method

In the final project, I plan to conduct the analysis via the statistical learning models, like random forest, linear discriminant analysis, and neural network. Then, I am going to compare their results by the classification accuracy. I am also interested in visualizing the final classification.