# Information Visualization Final Project - UCI Life Sciences Data

## Data description

#### Forest type mapping Data Set

Forest type mapping Dataset is from the UCI Machine Learning Repository(https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping ). The source of this dataset was collected in a forested area in Ibaraki Prefecture(36◦ 57′ N, 140◦ 38′ E), Japan, where contains mainly Cryptomeria japonica (Sugi, or Japanese Cedar) planted forest, Chamaecyparis obtusa (Hinoki, or Japanese Cypress) planted forest and mixed deciduous broadleaf natural forest. The total area where the data was collected is approximately 13 km × 12 km. Data was collected by 15 m spatial resolution multispectral Advanced Spaceborne Thermal Emission and Reflection Radiometer(ASTER) and removed the effects of image perspective(tilt) and relief (terrain) effects by Orthorectification. Only the information in three bands of the images was remained in the dataset: green (0.52–0.60 μm), red (0.63–0.69 μm) and near-infrared (NIR) (0.76–0.86 μm). (Johnson, Tateishi, & Xie, 2012)

This data set contains two parts: training and testing. the trainning set has 199 instances, testing data has 325 instances. Each dataset has 27 attributes. Since this is a dataset used to solve classification problems, the first column is the class of each instance: 's', 'h', 'd' and 'o', represent 'Sugi forest', 'Hinoki forest', 'Mixed deciduous' forest' and 'Other non-forest land respectively. From the second column to the tenth column(b1 to b9 ) are the ASTER image of three bands(green, red, and near infrared wavelengths) for three dates (Sept. 26, 2010; March 19, 2011; May 08, 2011). In order to minimize classification errors, the image data was predicted by using inverse distance weighting (IDW) interpolation method. The idea of this method is that the geographic data sets has spatial autocorrelation (O’Sullivan and Unwin 2003), which means that two areas that located close to each other share the similar natrure characteristics. In the dataset, column 11 to column 28 are the information created by this idea. Column 11 to 19(pred*minus*obs*S*b1 - pred*minus*obs*S*b9) are the predicted value minus actual spectral values for the 's' class(b1 -b9); column 20 to 28(pred*minus*obs*S*b1 - pred*minus*obs*S*b9) are the predicted value minus actual spectral values for the 's' class(b1 -b9);

* List of columns:

class, b1, b2, b3, b4, b5, b6, b7, b8, b9, pred\_minus\_obs\_H\_b1, pred\_minus\_obs\_H\_b2, pred\_minus\_obs\_H\_b3, pred\_minus\_obs\_H\_b4, pred\_minus\_obs\_H\_b5, pred\_minus\_obs\_H\_b6, pred\_minus\_obs\_H\_b7, pred\_minus\_obs\_H\_b8, pred\_minus\_obs\_H\_b9, pred\_minus\_obs\_S\_b1, pred\_minus\_obs\_S\_b2, pred\_minus\_obs\_S\_b3, pred\_minus\_obs\_S\_b4, pred\_minus\_obs\_S\_b5, pred\_minus\_obs\_S\_b6, pred\_minus\_obs\_S\_b7, pred\_minus\_obs\_S\_b8, pred\_minus\_obs\_S\_b9

### 

## Visualizations

## Observations of relationships

## Appendix

## Reference

1. Johnson, B., Tateishi, R., Xie, Z., 2012. Using geographically-weighted variables for image classification. Remote Sensing Letters, 3 (6), 491-499.