

Occupancy Detection

MT7038

Anton Strähle & Max Sjödin

Fall 2020

The occupancy status of a room was observed for a few days. Snapshots of the features below were taken every minute.

- ▶ Features
 - ▷ Temperature
 - ▷ CO2
 - ▷ Humidity
 - ▷ HumidityRatio
 - ▷ Light
- ▶ Response
 - ▷ Occupancy
 - ▷ Occupied
 - ▷ Unoccupied

Brief Exploration

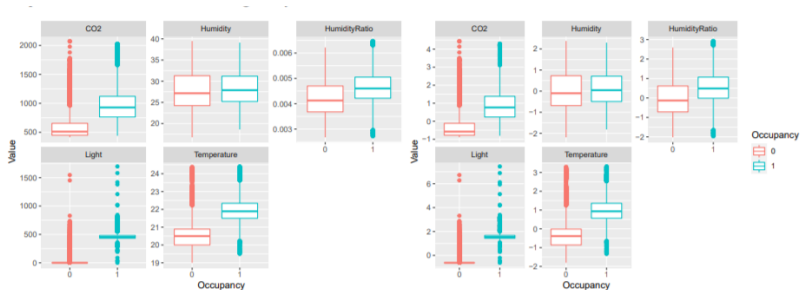


Figure: Boxplots of Features: Standardized and unstandardize

- ▶ Unbalanced data set

- ▷ Many more unoccupied data points than occupied

As the data set consists of minutely snapshots the sets cannot be combined and then resampled as we might have almost identical data points in all three sets.

Our solution to this problem was to upsample the Occupied class in both the training and validation sets so that we had an even split in both.

- ▶ SVM
 - ▷ Linear, Radial & Polynomial
- ▶ Logistic Regression
 - ▷ Regular, Boosted & Weighted

Methodology

SVM

Why? ▷ Good for classification and should generalize well with low costs

How? ▷ Using the package **e1071** and the function **svm**

▷ Linear, polynomial and radial kernels

▷ Coarse-to-fine parameter search

Kernel	Cost	TestAccuracy
Linear	0.00013	0.83940
Radial	0.00100	0.83490
Polynomial Degree 4	0.00004	0.83752

Figure: SVM Accuracies

Methodology

SVM

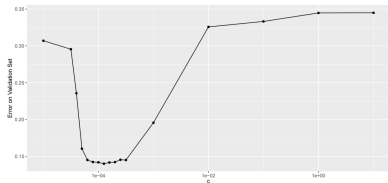


Figure: Linear

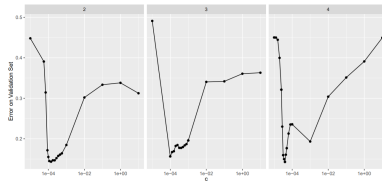


Figure: Polynomial

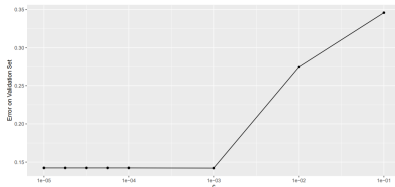


Figure: Radial

	Linear		Radial		Polynomial	
	0	1	0	1	0	1
0	1409	144	1376	124	1391	137
1	284	828	317	848	302	835

Figure: Confusion Matrices

Kernel	Cost	TestAccuracy
Linear	0.00013	0.83940
Radial	0.00100	0.83490
Polynomial Degree 4	0.00004	0.83752

Figure: SVM Accuracies

Methodology

Logistic regression

- Why? ▷ Good for classification and can be further optimized with boosting algorithms and regularization
- How? ▷ Regular GLM, Boosting, Lasso and Ridge
- ▷ Find optimal λ manually using validation set

Method	Test Accuracy
Logit	0.86979
Boosting	0.85854
Lasso	0.85516
Ridge	0.84390

Figure: Logistic Regression Accuracies

Discussion

Improvements

- ▷ Good for classification and should generalize well with low costs

How? ▷ Using the package **e1071** and the function **svm**

- ▷ Linear, polynomial and radial kernels
- ▷ Coarse-to-fine parameter search

Kernel	Cost	TestAccuracy
Linear	0.00013	0.83940
Radial	0.00100	0.83490
Polynomial Degree 4	0.00004	0.83752

Figure: SVM Accuracies