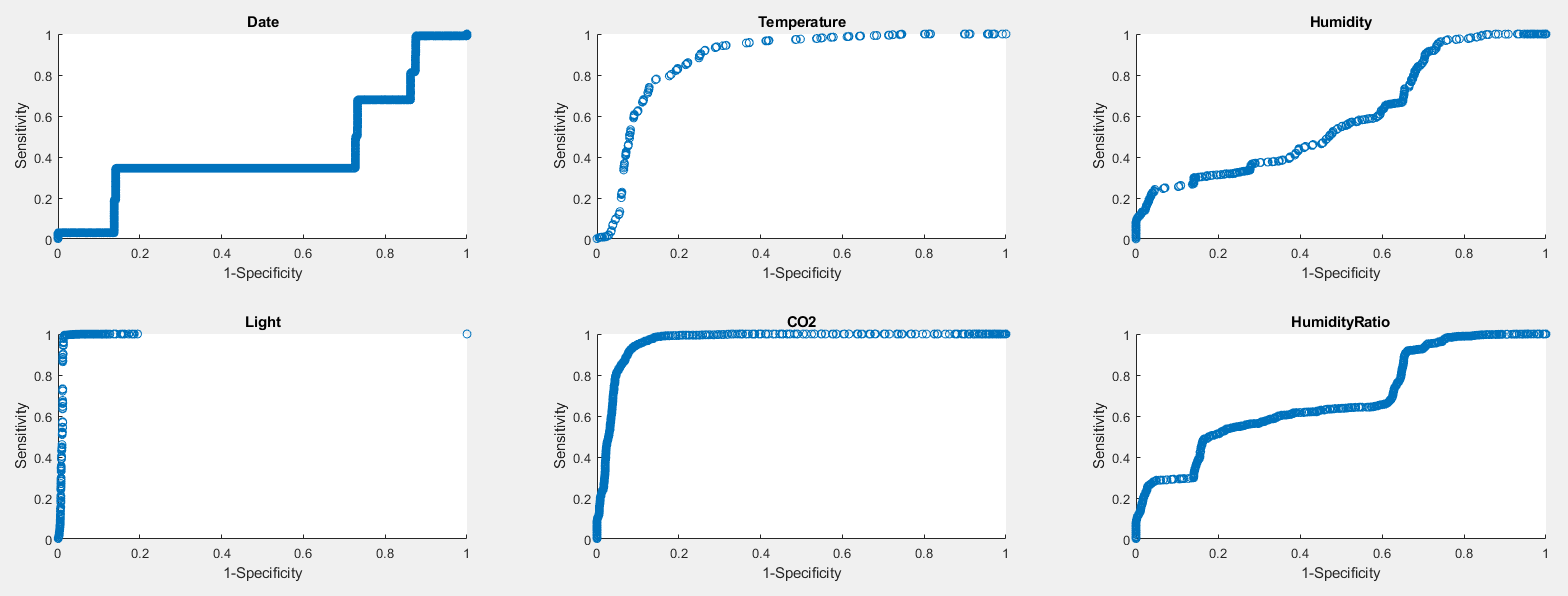
The K-fold error and AUC, overall optimal threshold, the error on unseen data given that threshold for each fold, can be seen in figure 1. According to the data obtained, “Light”, which resulted in an error of 0.02138837 on unseen data, is the best classifier. I came to this conclusion because it resulted in the lowest error on unseen data. This conclusion is ironic since the testing error was not the lowest of the other classifiers.

The ROC curves can be seen in figure 2 (AUC in figure 1). According to the curves, light performs the best on both the test data (0.1227998 error rate) and the unseen data (0.02138837 error rate). This occurs because out of all the other classifiers, the occupancy of the room is most closely related to the light. Therefore, this classifier is able to generalize really well. If the classifier had worked really well on the training data but poorly on the unseen data, we would have been experiencing overfitting. Overfitting occurs when you train a classifier using noisy data. In our homework, date would be an example of noise. As you can see from the ROC curve, the date and occupancy of the room are not closely related. Actually, we would have been better off guessing in this scenario.

The error on unseen data is satisfactory; therefore, by using light as the classifier, we were able to generalize really well. I would be interested in experimenting with testing additional classes or combining classes. Training with several classes that are both closely related to whether the room is occupied, would result in the classifier being better at generalizing. On the other hand, as discussed in class, creating a classifier which uses several classes that are not relevant to the occupancy would only add more “noise” and result in worse results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Average K-Fold Error** | **Average K-Fold AUC** | **Error on unseen data** | **Optimal Threshold** |
| **Date** | **0.20589439** | **0.43071184** | **0.36472795** | **2/10/2015 8:38** |
| **Temperature** | **0.15500068** | **0.87595747** | **0.14634146** | **21.5693181** |
| **Humidity** | **0.19306863** | **0.5912116** | **0.36472795** | **36.9040343** |
| **Light** | **0.01227998** | **0.99041837** | **0.02138837** | **364.3866085** |
| **CO2** | **0.07722745** | **0.96350959** | **0.13058161** | **742.2250031** |
| **HumidityRatio** | **0.18024287** | **0.68057644** | **0.28818011** | **0.00509599** |

**Figure 1. K-Fold Results**



**Figure 2. K-Fold ROC Curves**