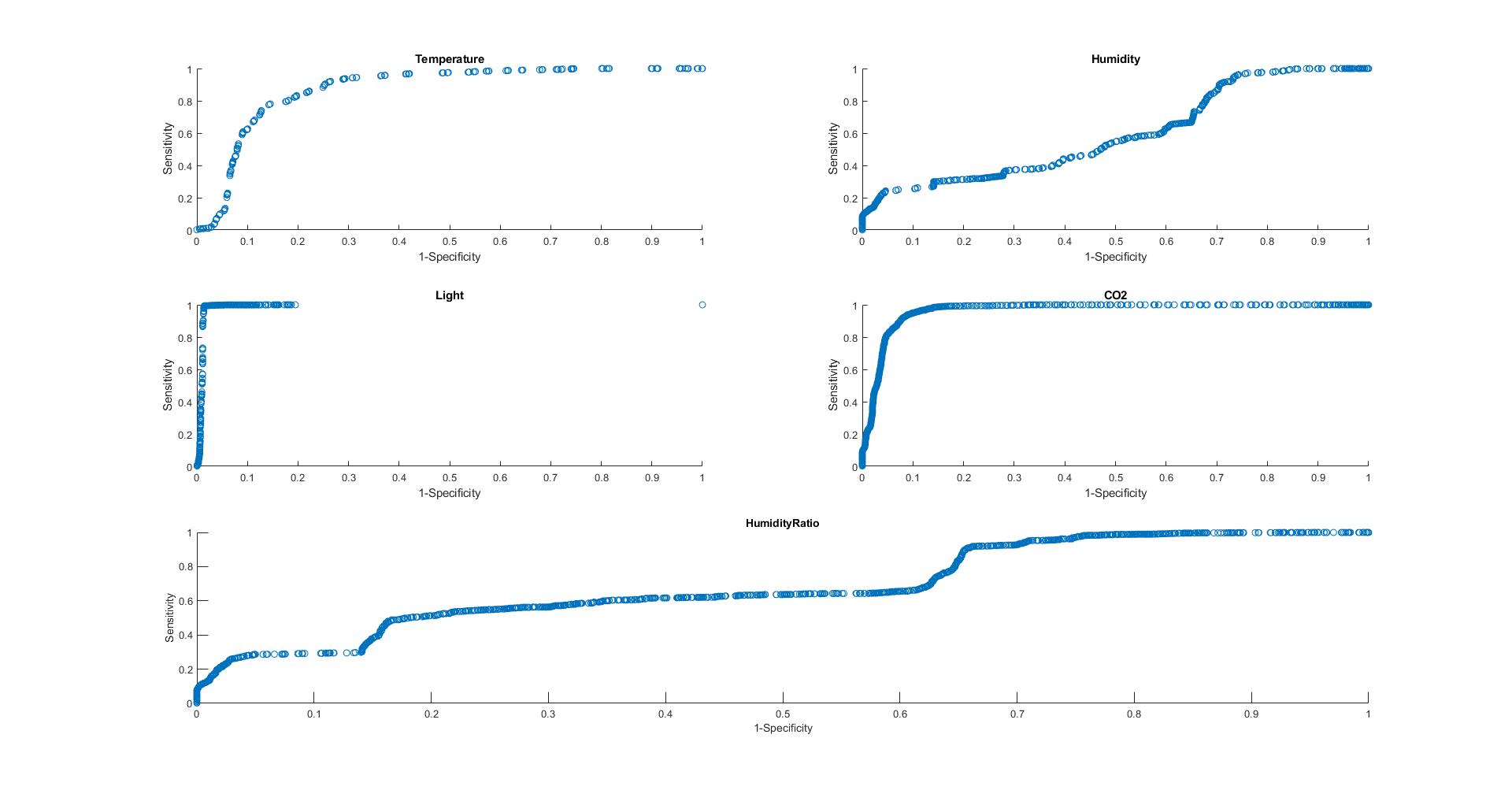
1. Compare the classifiers in terms of error measures shown on slide 24 (“Measuring errors”). Compute ROC and AUC for each classifier. Determine the best classifier and corresponding error.
2. Now use data from “datatest.txt” and apply your best classification model. Assess classification errors. Provide conclusions about the generalization capabilities of the developed classification model.

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, temperature, which resulted in an error of 0.14634146 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 the test data but does not do so on the unseen data. This can be attributed to overfitting, which occurs when you train a classifier using noisy data. On the other hand, the temperature classifier had an average performance on the training data but ended performing better than the other classifiers on the unseen data. This tells us that the occupancy has a closer relationship with the temperature class than the others.

The temperature was able to generalize fairly well. The error on unseen data is satisfactory. I would be interested in experimenting with testing additional classes or combining classes. Training with several classes that are both closely related to whether or not the room is occupied, would result in the classifier being better at generalizing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Average K-Fold Error** | **Average K-Fold AUC** | **Error on unseen data** | **Optimal Threshold** |
| Temperature | 0.15500068 | 0.87595747 | 0.14634146 | 21.5693181 |
| Humidity | 0.19306863 | 0.5912116 | 0.36472795 | 36.9040343 |
| Light | 0.01227998 | 0.99041837 | 0.36472795 | 364.3866085 |
| CO2 | 0.07722745 | 0.96350959 | 0.36472795 | 742.2250031 |
| HumidityRatio | 0.18024287 | 0.68057644 | 0.63527205 | 0.00509599 |

**Figure 1.**

**Figure 2.**