Again, we apply SVM with whole Seattle data as training set and test on NYC. As shown in Table 1, the Type I error is significantly higher as well as the Type II error. And the Type II error is much less than the Type I error. One of possible causes is that the dataset for NYC is imbalanced which have more days with precipitation, 238 non precipitation days and 126 precipitation days.

**Confusion Matrix for New York Data in SVM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | True condition | |  |
|  |  | Non-rain | Rain |  |
| Prediction | Non-rain | 96 | 45 |
| Rain | 142 | 81 |
|  | Error rate | 0.596 | 0.357 | 0.5137 |
|  | Type I | Type II | Overall |

Table 1