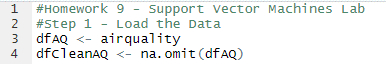
Jeffrey Chao

IST 687 – Homework 9

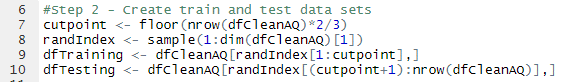
Date Due: 03/17/2020

Date Submitted: 03/17/2020

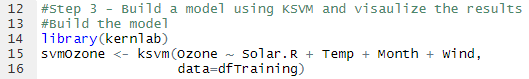
**Step 1: Load the data**

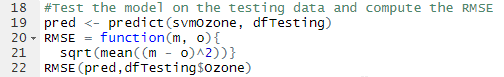


**Step 2: Create train and test data sets**

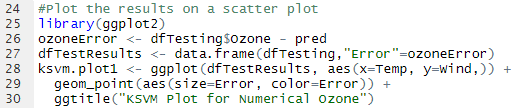


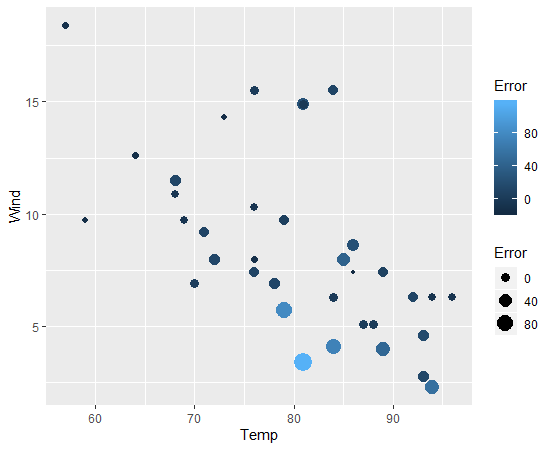
**Step 3: Build a Model using KSVM & visualize the results**

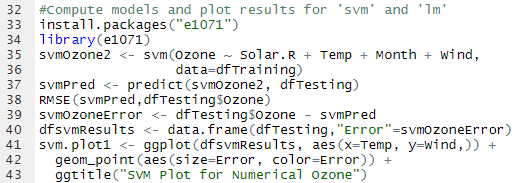


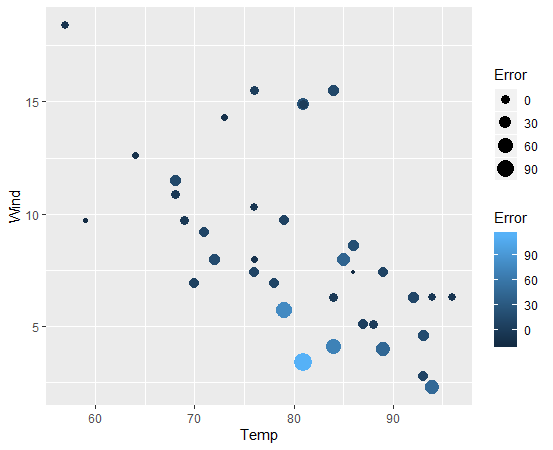


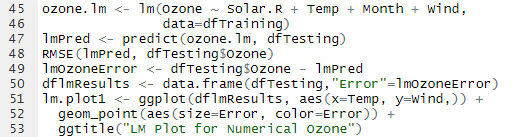


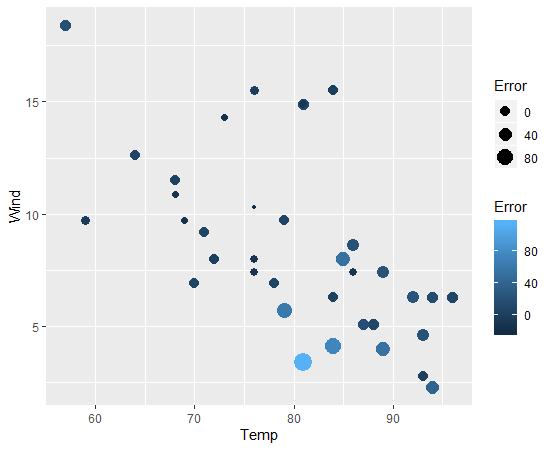












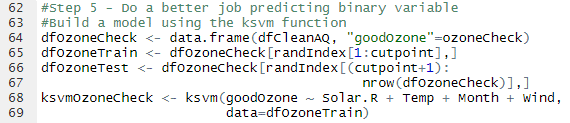


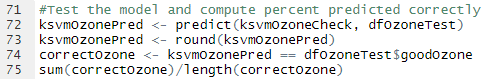


**Step 4: Create a ‘goodOzone’ variable**

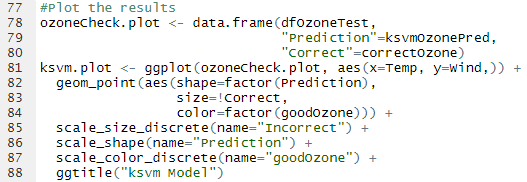


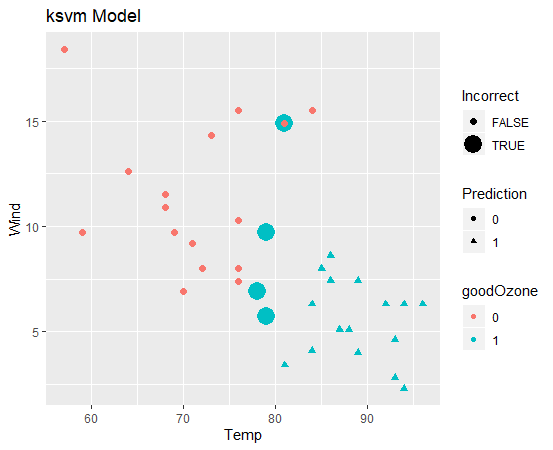
**Step 5: See if we can do a better job predicting ‘good’ and ‘bad’ days**

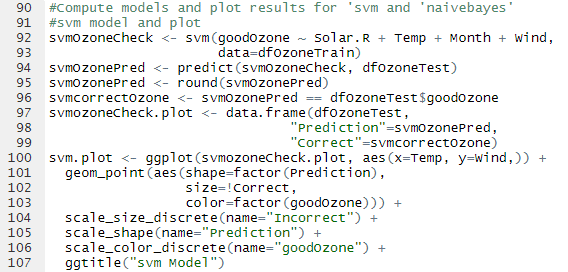


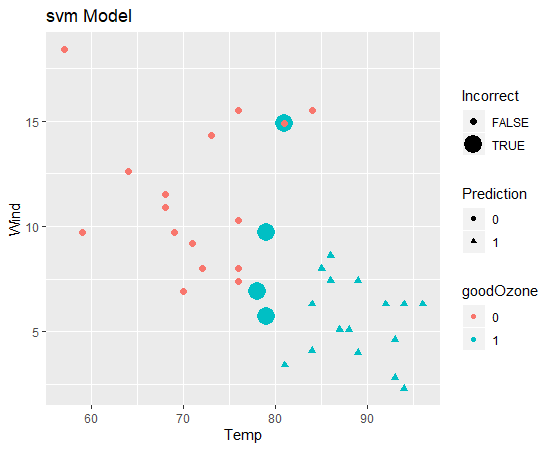


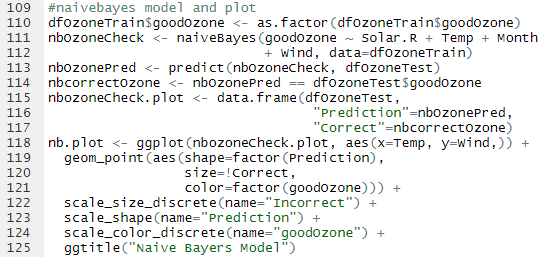


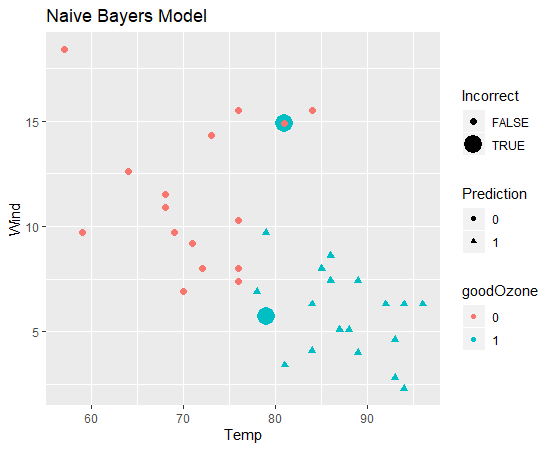




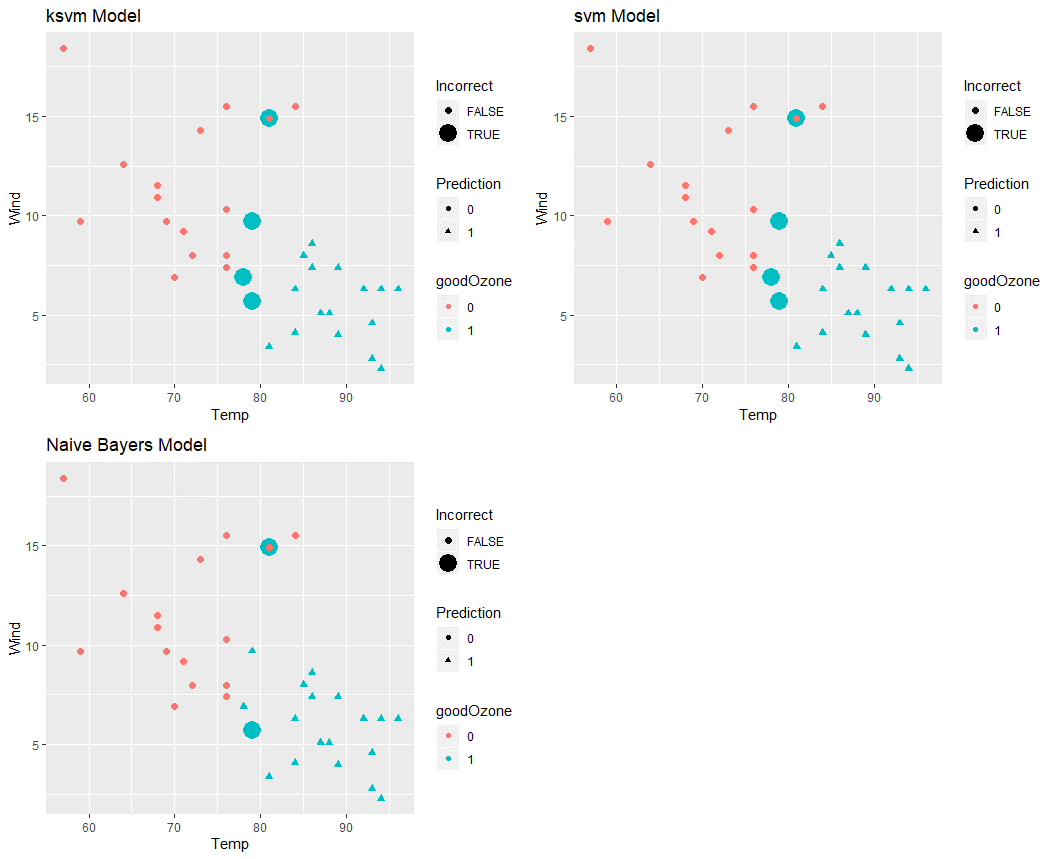












**Step 6: Which are the best Models for this data?**

Comparing the numerical Ozone models to the goodOzone models, it is very clear that the prediction power is much more accurate if the predicted variable is discrete (rather than continuous). On a further breakdown, for the training and testing data used in this instance, it looks like the Naïve Bayers model greatly outperforms either SVM or KSVM models by only have 2 incorrect prediction compare to the 4 incorrect in the other models. However, since this data set was so limited (only 111 observations total, <100 rows for training and <40 rows for testing) this conclusion may change greatly if tested on a different set of data.