# Chapter 7 Homework

Cody Frisby

10/14/2017

The predicted classes and the actual classes are shown in the below table using the proportion of each class label for comparison. If the **score** data were a simple random sample from the same populaiton as our **train** data then it would seem reasonable to expect similar proportions for the 4 classes. This appears to not be thes case. Our LDA model is predicting a lot more cars than what we observed.

	Car	SUV	Truck	Van
Observed	0.589	0.111	0.189	0.111
Predicted	0.889	0.000	0.067	0.044

Visualizing the predicted classes for the *Score* data set in the first two LDA dimensions yields the following plot.
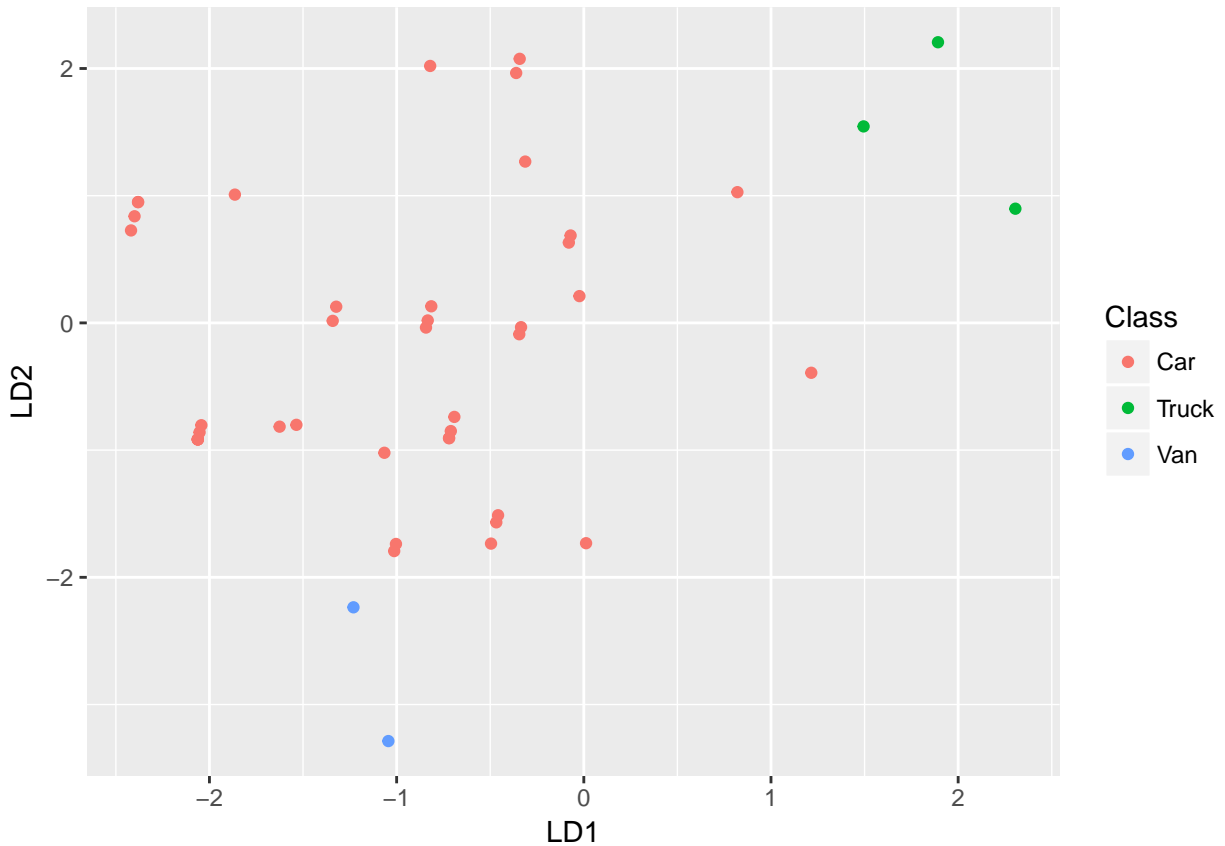
It appears that the predicted classes aren't so great for the **Score** dataset, but alas, it is what it is. I do not have the actual class variables so I cannot test how well the model predicts.

Fitting all three models and then using the models to predict the classes for the **scoring** data we get comparision results that are shown in the table below. It can been seen that all the models are not predicting the same.

	Car	SUV	Truck	Van
LDA	40	0	3	2
KNN	38	2	4	1
NaiveBayes	28	0	0	17

Since I don't know the classes of the **Scoring** data there is no way to know how well my models are working and which is best. It would be better to have a test set in addition to the training set so that we can evaluate each model with data that was NOT used to train it. Anyways, as can be seen from the above table, *LDA* and *KNN* predict the same number for $class = Car$. They are almost identical except for a few differences between the classes of SUV and Van. The *NaiveBayes* model is the most different, where it is predicting a lot fewer of $class = Car$ than the above two. It also is predicting a lot more of $class = Van$ than the other two models. The linear discriminant model and the K nearest neighbor model almost agree perfectly where $k = 1$ and $l = 0$ as parameters to the KNN model.

Since the **score** data is small enough, we could examine the observations while included the predicted class for each element from each model. Where the last three columns agree, we have agreement between all three models for that observation. But what proportion of our **score** data has agreement between all three models. It appears that the only time that all models aggree is when they all predict $class = car$. This kind of makes sense, since if someone doesn't yet have a vehicle then their first vehicle is most likely to be a car rather than a more expensive SUV, truck, or van. The proporiton of **score** predictions that agree between all three models is 0.48889.

Score data with added model predictions for reference

Gender	Age	Marital_Status	Employment	Housing	ldaClass	knnClass	nbClass	Agree
0	32	0	1	2	Car	Truck	Car	FALSE
1	22	0	2	1	Car	Car	Car	TRUE
1	21	0	2	1	Car	Car	Car	TRUE
1	20	1	0	1	Car	Car	Van	FALSE
0	18	1	0	1	Car	Car	Van	FALSE
0	19	1	0	1	Car	Car	Van	FALSE
0	18	0	0	1	Car	Car	Car	TRUE
0	19	0	0	1	Car	Car	Car	TRUE
1	21	2	0	1	Car	Car	Car	TRUE
0	26	3	1	1	Van	Car	Car	FALSE
1	27	0	1	2	Car	Car	Car	TRUE
0	25	1	2	1	Car	Van	Van	FALSE
1	24	1	2	1	Car	Car	Van	FALSE
1	25	1	1	1	Car	Car	Van	FALSE
0	26	2	0	1	Van	SUV	Car	FALSE
0	41	0	1	1	Car	Truck	Car	FALSE
1	44	0	1	1	Car	Car	Car	TRUE
1	63	1	3	0	Car	Truck	Van	FALSE
0	80	1	3	0	Car	Car	Van	FALSE
0	26	1	1	1	Car	SUV	Van	FALSE
1	25	1	2	1	Car	Car	Van	FALSE
1	22	1	0	1	Car	Car	Van	FALSE
0	18	0	0	1	Car	Car	Car	TRUE
0	21	0	0	1	Car	Car	Car	TRUE
0	22	1	1	1	Car	Car	Van	FALSE
1	15	0	0	0	Car	Car	Car	TRUE
1	17	0	0	0	Car	Car	Car	TRUE
0	16	0	0	0	Car	Car	Car	TRUE

Gender	Age	Marital_Status	Employment	Housing	ldaClass	knnClass	nbClass	Agree
0	21	0	1	0	Car	Car	Car	TRUE
1	90	3	3	2	Truck	Car	Car	FALSE
0	77	1	3	2	Truck	Car	Van	FALSE
1	42	0	1	1	Car	Truck	Car	FALSE
0	25	1	1	1	Car	Car	Van	FALSE
1	23	1	1	1	Car	Car	Van	FALSE
1	19	0	0	0	Car	Car	Car	TRUE
0	17	0	0	0	Car	Car	Car	TRUE
0	16	0	0	0	Car	Car	Car	TRUE
0	16	0	0	0	Car	Car	Car	TRUE
1	23	0	1	0	Car	Car	Car	TRUE
0	20	0	2	0	Car	Car	Car	TRUE
0	18	0	0	0	Car	Car	Car	TRUE
1	70	1	3	2	Truck	Car	Van	FALSE
1	22	1	1	1	Car	Car	Van	FALSE
0	16	0	0	0	Car	Car	Car	TRUE
1	19	0	0	0	Car	Car	Car	TRUE