Machine Learning – Assignment #1 – Task 3
U1425964

The most predictive attribute for classification of plants or animals in the given dataset is TNA. The following two statistics were used to come to this conclusion.

Regression coefficients

Since the attributes are standardised, the absolute value of the coefficients of the predictors should tell us how much each of the attributes affect the final result. Comparing coefficients from fitting the ClassA dataset using both ridge and lasso regression gave consistent results. TNA had a coefficient around 0.27 while the others were far below these figures.

Change in AUC with Knn

AUC values when each attribute is dropped one at a time with Knn was compared to confirm the findings from above. The AUC results is given below.

```
TNA 0.57755656403
c1 0.78244229737
c2 0.716384781795
c3 0.707689292279
m1 0.725499194847
m2 0.707505814994
m3 0.757363002001
n1 0.70984669562
n2 0.759209892809
n3 0.668285756112
p1 0.730037248491
p2 0.705652092584
p3 0.730646399584
```

Clearly, absence of TNA caused the biggest drop in AUC implying that they have a higher predictive power in this dataset.