

Results

Anirudh

December 9, 2022

Redshift Range

Most of the Associated spectra we have has a redshift range of $0.02 < z < 0.25$ as the most of the associated absorbers we had are from Maccagni et al. 2017 and Gereb et al. 2015, though we have 10 Associated spectra with redshift $z > 0.25$ reaching till $z = 1.2$. Whereas our intervening spectra have (mostly) $z > 0.25$ (There are very few, around 3-4 with $z < 0.25$).

Busy Function

I will attach the image generated from busy function, with the free parameters mentioned in the figure 1.

$$B_1(x) = \frac{a}{4} (\text{erf}[b_1\{w + x - x_e\}] + 1) (\text{erf}[b_2\{w - x + x_e\}] + 1) (c|x - x_p|^n + 1) \quad (1)$$

Also note that the busyfit software also gives parameters like line centroid, half width, 20% width, Peak flux and integrated flux for the spectra.

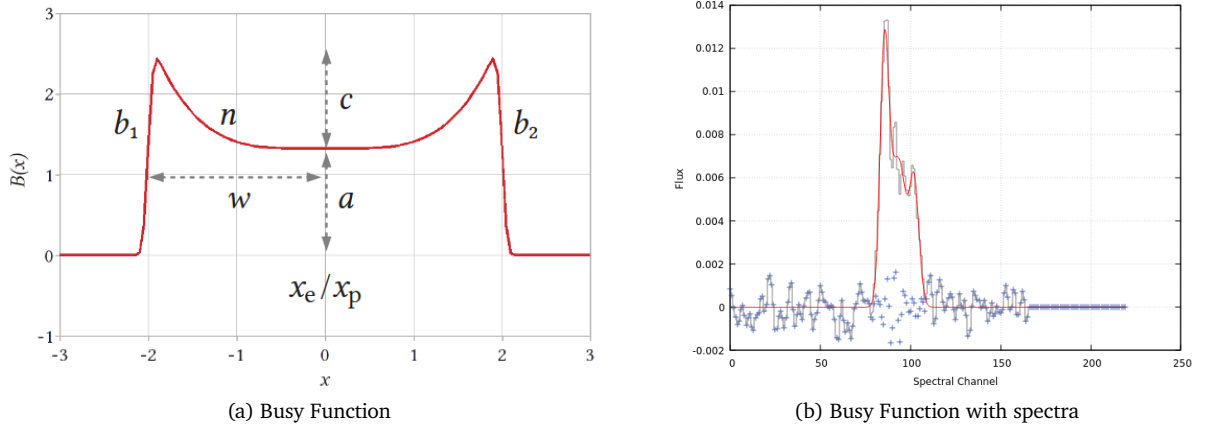


Figure 1: Busy function plots with parameters and spectra respectively. (The spectra used is an Associated spectra)

ROC Curves along with ROC AUC, Accuracy and Precision results with the removal of Redshift dependant parameters

I have removed all the redshift dependant parameters like line widths and centroid position in the spectrum were removed and observed the following results

i

Info: These results are obtained using SMOTE to overcome class imbalance. ROC AUC is the area under the ROC curve. Also it should be noted that these are all average value of 30 folds i.e The data is split into 10 sets three times differently.

ML Model	ROC AUC	Average Accuracy	Average Precision
Random Forest	0.91	86.8%	81.8%
KNN	0.72	79%	60.63%
Decision Tree	0.799	84.21%	56.6%
Logistic Regression	0.78	85.2%	73%
SVM	0.698	84.15%	64.6%

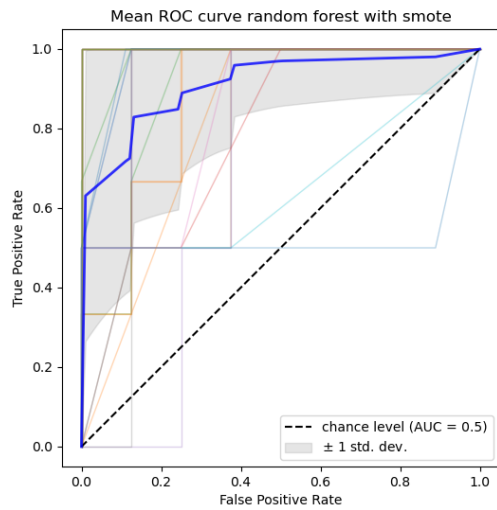
Table 1: Results for Machine Learning models after removing redshift dependant parameters.

I am also attaching the ROC plots for all the Machine Learning Models in Figure 2 (Next Page)

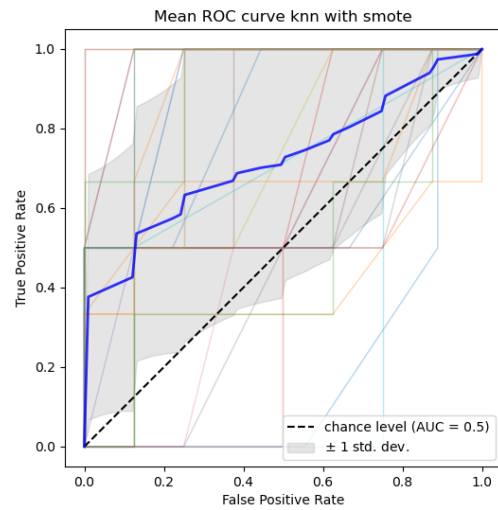
i

Info: In the below plots, you will notice a lot of curves, the mean ROC curve for all the 30 folds is the thick blue curve. The other curves (thin) are for each fold validation, I haven't mentioned each legend as 30 legends won't fit. The plots also contain $\pm 1\sigma$ (Standard Deviation).

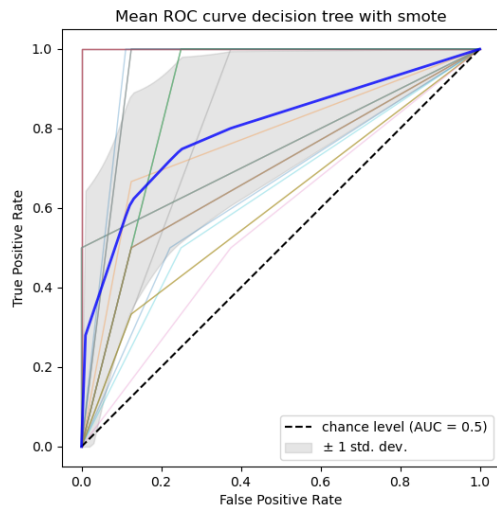
From the figures, we can note a decrease in ROC AUC, Accuracy and Precision. This can explained that the parameters which were redshift dependant played a major role in differentating the spectra into associated and intervening as the spectra does not have much overlap over the redshift range. Another reason can also be because we removed the differentiating factor (width) for associated and intervening spectra, as intervening spectra do have narrow line width compared to associated spectra spectra (Gupta et al. 2009 and Holt et al. 2008). Though width plays an important role, it still has high redshift dependance, either way we have a high redshift involment in the parameters.



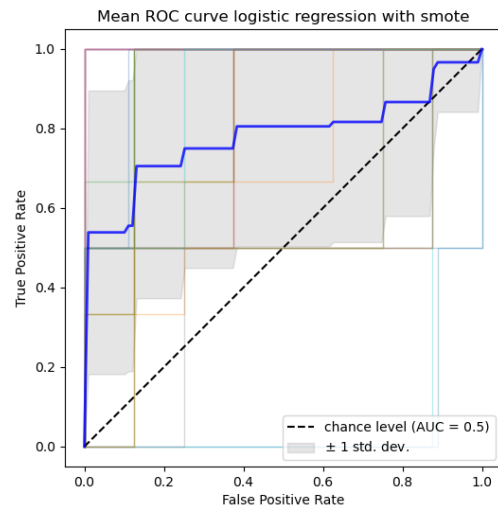
(a) Random Forest



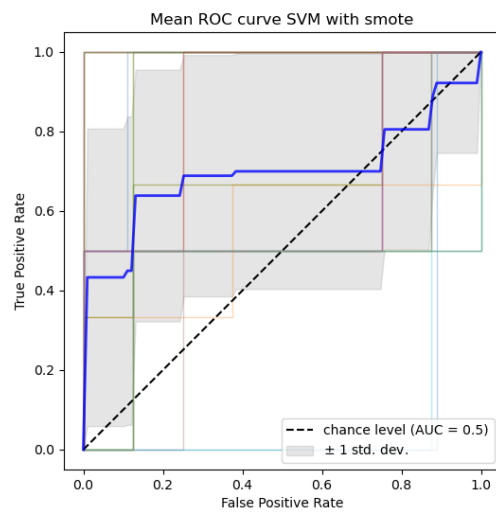
(b) KNN



(c) Decision Tree



(d) Logistic Regression



(e) SVM

Figure 2: Mean ROC curve plots for all models