1 Introduction

Gamma-ray bursts (GRBs) are extremely energetic explosions that are observed in cosmological distances from all over the Universe isotropically. With luminosities of $10^{51\text{-}52}$ergs/s, they are the most luminous objects in the sky (Piran 2005). Despite being discovered in 1973 (Klebesadel et al. 1973), the exact physical mechanism that produces GRBs are still unknown, making them one of the most intriguing puzzles that modern day high-energy astrophysics have to offer. The most popular scenario proposed for explaining GRBs is the so called Fireball Model (Paczynski 1986, Goodman 1986, Rees Meszaros 1994). In this picture, it is a stellar mass black hole or a highly magnetized neutron star which is the remains of the death of a supermasive star that powers the GRBs alongside the mergers of two compact objects that accounts for the bursts of shorter durations (Woosley 1993, Paczynski 1986). Both of these events result in bipolar collimated jets and hence a relativistically expanding fireball consisting of photons, electrons, positrons and a small amount of baryons. The emission observed results from the interaction of layers of different velocities, resulting in shocks that accelerate particles that later cool down and emit via different processes such as synchrotron, Synchrotron Self Compton (SSC) and Inverse Compton (IC). A further prediction of the Fireball model is the existence of a highly thermal region deep inside the flow, from where a thermal emission (blackbody) is expected, intertwined with the non-thermal radiation occuring later in the flow. Featuring both thermal and non-thermal components, the GRB prompt spectra have been interpreted with many different spectral models (see Burgess et al. 2011 and Ryde 2005). The most traditionally used model is the Band function, which is an empirical model that simply consists of a smoothly broken powerlaw (Band et al. 1993). This function gives three parameters that have proved quite useful in classifying different GRB spectra, namely, the low-energy slope α, the high-energy slope β and the peak energy, E_{pk}. The values of these parameters, especially that of αs, help interpreting GRB spectral shapes in terms of different physical radiation processes given the constraints they put on the parameters. It is, therefore, of great importance to learn more about whether the observations at hand show any trends in this parameter space or not and if so, how they group and at which values. The aim of this machine learning project is to find out the answer to this question by applying a number of different clustering algorithms to a fairly large data sample obtained from Axelsson Borgonova 2014. This data set includes the Band function fit parameters of 1145 GRBs observed from 2008 to 2015 with their fluence, flux and duration information and hence, is a very convenient tool to study the possible clustering in the parameter space defined by the Band function.

2 Machine learning approach

Question that must be asked first when deciding to imply machine learning algorithms to a certain problem is: what is the most suitable approach for given number in my data sample. If we look at the "cheat -sheet" on the scikit-learn webpage we see that if we have less than 50 samples in our data sample we should get first more data before proceeding with the analysis. Our data sample consists of 1144 GRBs with 16 different properties. Some of them are: T_{90} (time of in which a GRB emits from 5% of its total measured counts to 95%), α and β spectral indices and E_{peak} which is the peak energy of the prompt GRB emission. Since we have labeled data we could either do a classification analysis or a clustering analysis. We decided for a clustering analysis to inspect if there exist clusters in α and β spectral indices. Since we have idea of what we want to achieve this is also a supervised learning problem. If we were to have unlabeled data or to do analysis without any prior knowledge about the data set we would then employ methods of unsupervised learning.
We decided to use Mean Shift method to inspect how many categories we would get as well as to use KMeans, Gaussian Mixture Models (GMM) and Hierarchical Clustering to test our prediction for potential number of clusters.

2.1 Mean Shift

(Another) non-parametric method chosen was the Mean Shift. Since the search in this data sample is mainly focused on finding clusters without any specified shapes or distributions, non-parametric density estimation and grouping the data points around their closests respective means conceptually is a good starting point for understanding the degreee of clustering in the data. The mean shift algorithm was applied to our data frame in the csv format with ease by making use of the tools from Astropy, AstroML Sckitlearn, Numpy and matplotlib packages. The input had to be rescaled in line with the algorithm's requirements. In this task, the focus was the distribution of α values with respect to β values of all GRBs in our sample, since they are the main parameters that determine the shape of the spectra. There was a range of possibilities in choosing the bandwidth. The bandwidth estimated by the algorithm itself has given a value of 2.72, which, when applied to the data set, gave rather inconclusive results which can be seen in Fig1. The estimated bandwidth gave the bulk of the sample in one big cluster and some data points that can be considered rather as outliers are demonstrated as many tiny individual groups which does not help with understanding how the natural, smooth transition between different spectral shapes occur within our GRB sample. Trials with other bandwidths were made. As expected, larger bandwidths gave even lesser information about the clustering in our data set, which can be thought of as analogous to having a very low resolution while observing our data and a seeminlgly biased output. Therefore, smaller bandwidths were preffered but used with caution, since a small enough bandwidth would again diminish the information that can be gathered from the output of the algorithm by having a large variance. The

optimal bandwidth was found to be 0.6, for which the resulting clustering can be seen in Fig2. It can be seen that the mean shift algorithm concludes that most of our sample do not constitute densely populated clusters and hence remains as background while many smaller clusters are found and looking at the larger picture, at least 5 different larger locii for the clusterings can be identified.

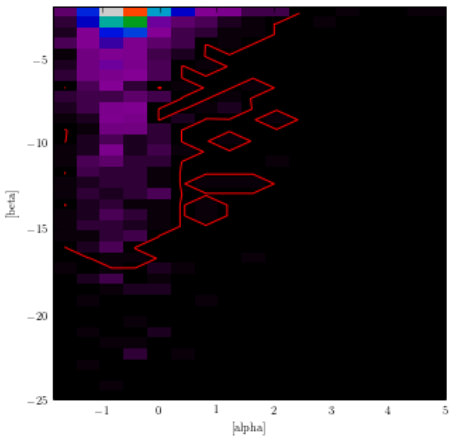

Figure 1: Clustering with the mean shift algorithm for bandwidth value 2.72.

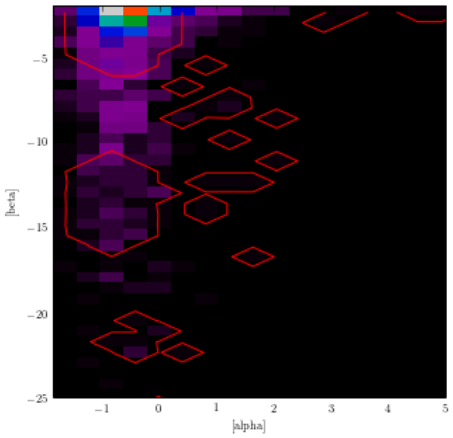

Figure 2: Clustering with the mean shift algorithm for bandwidth value 0.6.

2.2 Gaussian Mixture Model

For robustness, a parametric density estimation was also applied to the data set. Gaussian mixture model (GMM) was selected for this task since it was easy to use and quite descriptive with information criteria such as BIC and AIC giving the optimal number of clusters. As with the mean shift algorithm, α and β values from the data set were used, this time without the need of rescaling. The number of minimum cluster to describe the data is estimated to be 6 by the above mentioned information criteria implemented in the AstroML GMM algorithm. The results can be seen in Fig3 which represents the input, the plot of information criteria with respect to number of components and the final clustering, respectively. In Fig3, how these 6 clusters are located in the parameter space can be seen as well as the intersection of some clusters can be identified. It is seen that the results obtained from GMM clustering is very informative, especially when compared with other spectral parameters of the GRBs in the sample.

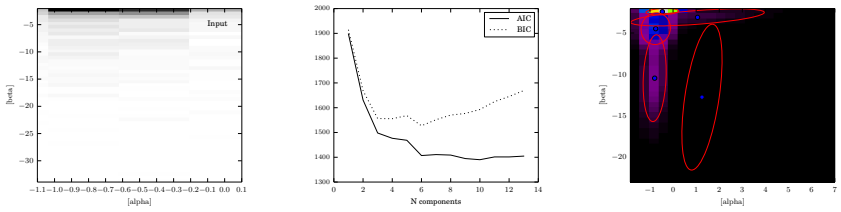

Figure 3: Clustering with the GMM algorithm.

2.3 KMeans

KMeans is the simplest method used from all of the above mentioned. It seeks partitioning of data points into N disjoint subsets C_n with each subset containing K_N points. It is encoded into scikit-learn and can be invoked by command KMeans. The command takes the number of possible clusters and data set and does clustering on it. In the Fig4 it can be seen result of KMeans algorith applied to our data set. The result recovered depends greatly on the number of clusters given as an input in the code. Here presented is the result for inital guess of 5 clusters and the algorithm recovers 3 of them (or 5 wheras 2 completely overlap).The alogrith was also tested with 4,6 and 10 clusters. When given input of 4 clusters algorithm recovers 2, for 6 recovers 3 and for 10 the plot looks messy and unreasonable.

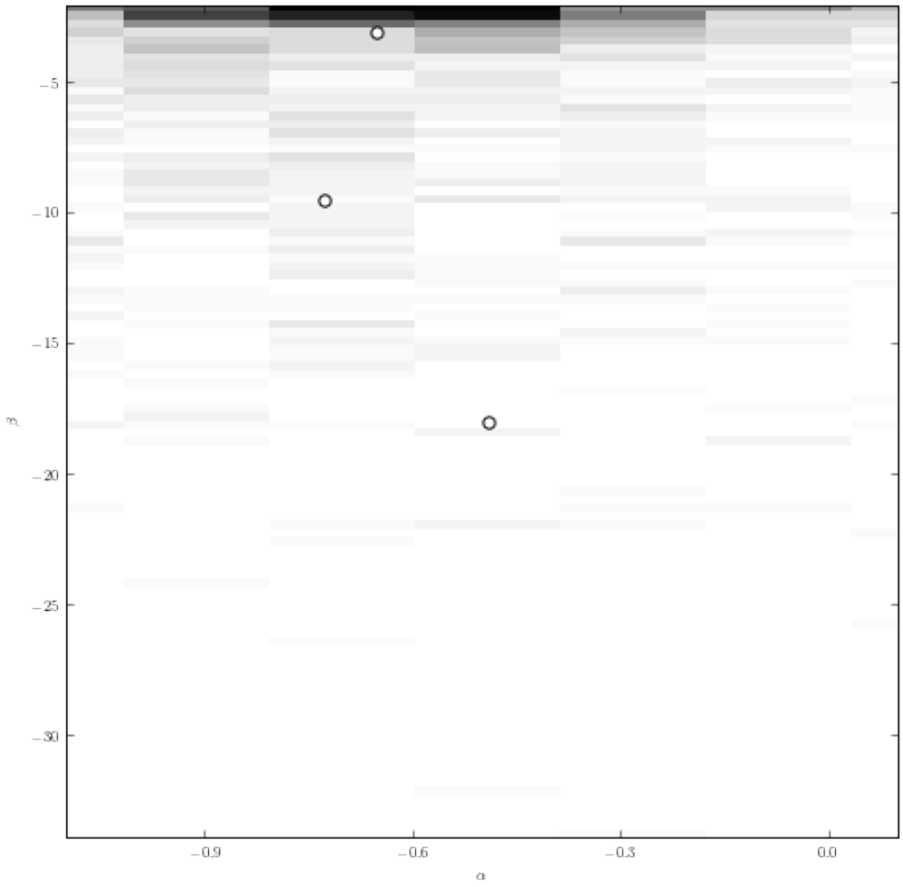

Figure 4: Clustering with KMeans algorithm

2.4 Hierarchical Clustering

This method, unlike previous, does not require input on number of clusters in data sample. It starts by dividing data into N clusters one for each data point in the data set. In the next step two clusters are merged and the resulting number of clusters is lowered by one and is N-1. This is then repeated until Nth partition contains one cluster. This process is visualized using a tree diagram or dendrogram. After applying Hierarchical Clustering function built in astroML we obtained the dendrogram plotted at Fig5. If we interpert darker areas as clusters with multiple brances then this algorithm recovers 2 clusters and all other branches are clusters with only one component in them. Since this method does not require as an input number of clusters it follows naturally the structure in the data sample merging clusters together.

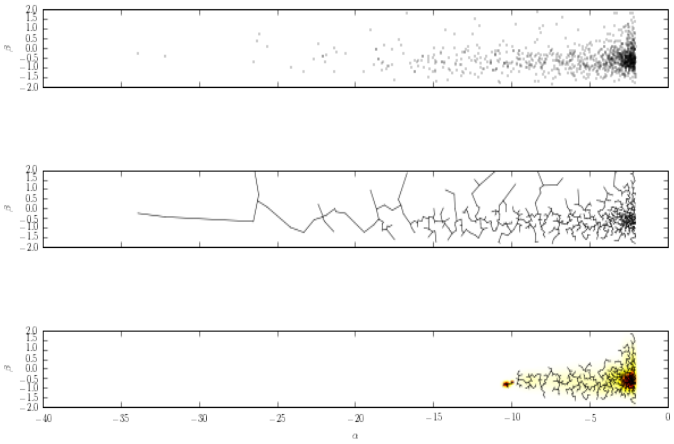

Figure 5: Hierarchical clustering on our GRB data set. On x-axis is plotted α and on y β spectral index

3 Results and discussion

A parametric and a non parametric clustering method was used to understand the nature of the distribution of the relation of low-energy slopes to high-energy slopes for GRBs fitted with a smoothly broken powerlaw specifically called a Band function. The non-parametric mean shift method gave many different clustering patterns dependent on the selection of bandwidth. Although this method gives some interesting insights to the data sample we have, it does not capture the smoothly changing nature of the parameters in each GRB spectrum that is compared, by taking some of the sample as mere background. The outcome of the optimal bandwidth for the mean shift algorithm could be used to determine the extremes in the sample and act, in a sense, as a gradient for assesing the direction of change and for what values of α and β that the maximal density of GRBs is achieved on the parameter space. The parametric GMM method, on the other hand, is much more suitable for the data sample at hand, for it makes the trial of different bandwidths or bins redundant by presenting the user with the choice of using BIC and AIC to asses the number of optimal clusters first, without the need of deciding on any binning that would affect the appearance of the data later on. GMM gives 6 clusters for our data, the three of which significantly overlap. This is very informative from our perspective, since it is known that the sample at hand is not distinctly separated and it is rather the purpose of this endeavour to find how different clusters that can be identified via spectral analysis on each GRB are related to each other on a larger scale. With the incorporation of other information such as the luminosity and the variability time scales to the clustering given by GMM, we are able to track the continous transition and outlying groups in our sample distinctively. While

for KMeans number of clusters that will be recovered depends on the initial guess for number of clusters. When given to look for 4 clusters it recovered 2, for initial guess of 5 and 6 clusters it recovered 3 and so on. On the other hand, Hierarchical clustering will not depend on the initial number for number of clusters, since it does not need that parameter in clustering. It has more potential to capture the underlying structure in the data sample.
Since these methods gave different results as a conclusion it can be said that either more objective clustering method is needed (such as Dirichlet processes) or clustering method is not applicable to our dataset and we need to look for structure in data set using some other methods, as classification. It is also important to stress that we did not do cross validation which could tell us much more about the process and how good it is in recovering the underlined structure.
These methods are useful in helping us getting a sneak peak into our data set but if used wrongly and not interpreted properly they will lead analysis into wrong direction. As with any automated analysis sanity check is need when employing supervised machine learning algorithms no matter how powerful they are.