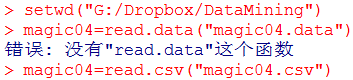
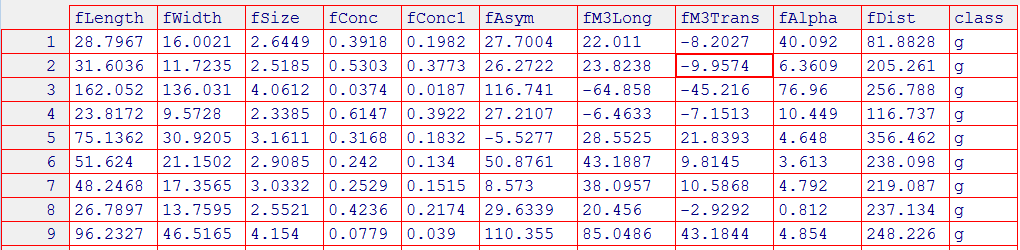
Data Mining HW2

Chengjun Yuan

cy3yb@virginia.edu

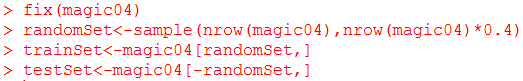
**1.**

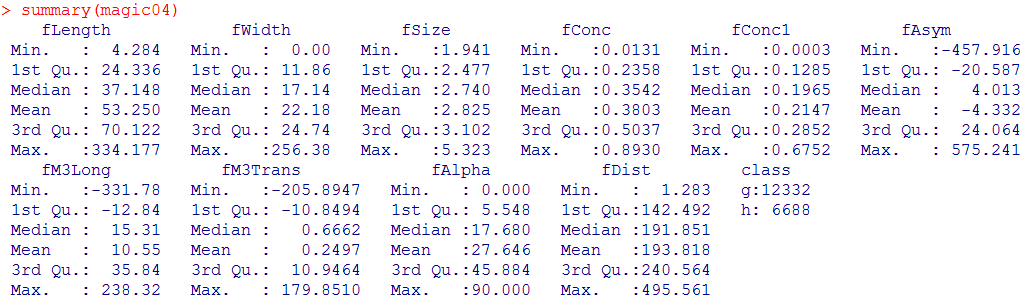


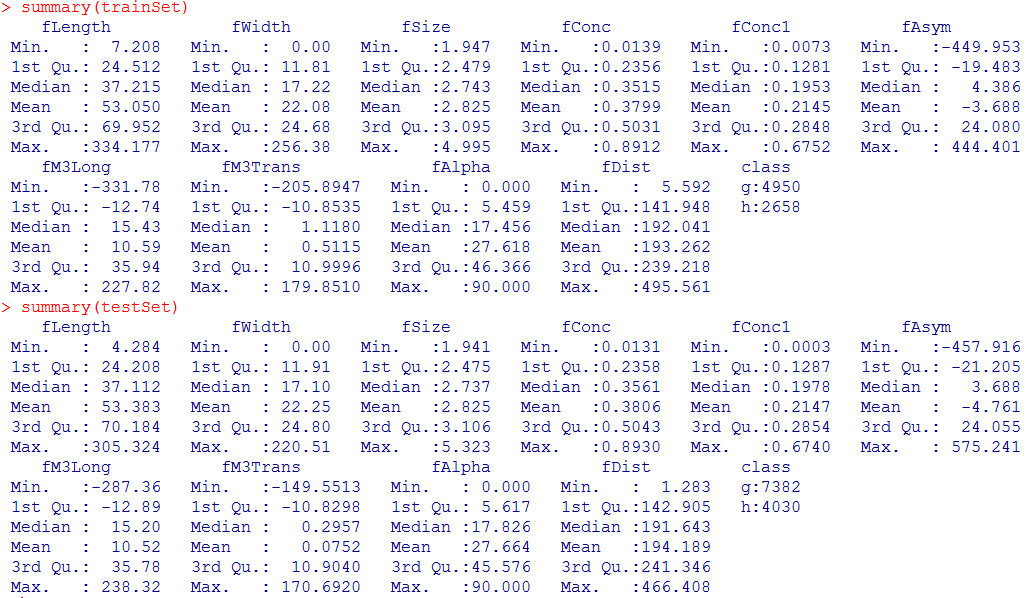


This data collects the image information of total 19020 rays. These rays come from two kinds of sources. One is from the primary gamma that is the **signal** we want. The other is the cosmic rays from the upper atmosphere which is the **background** we do not want. Therefore, we need to use the image information to discriminate statistically those rays by the primary gammas (**Signal**) from the cosmic rays by atmosphere (**Background**).

**2. Randomize the entire data into training set and testing set**

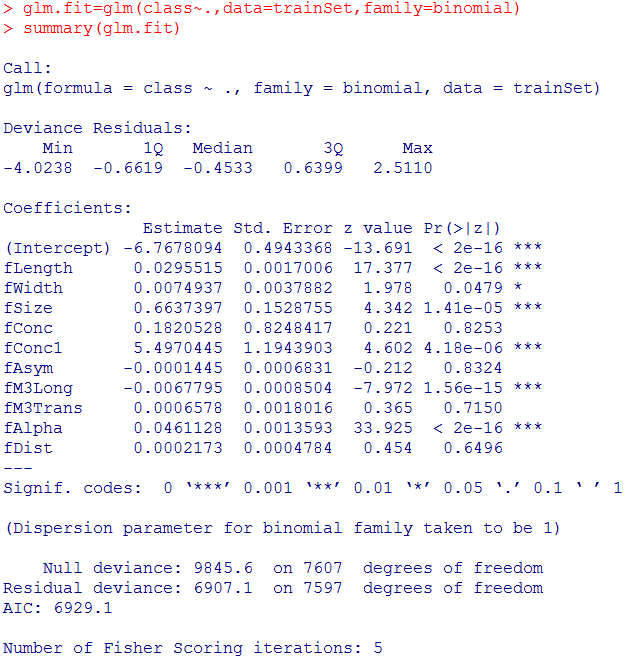


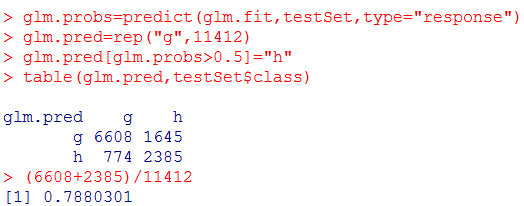




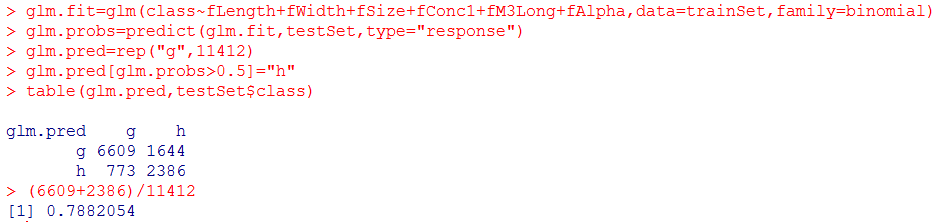
Here, we randomly choose 40% of the whole data set as the training data set, and the rest is the testing data set. So, there are 7608 observations in training set and 11412 ones in testing set. From the summary of these three data set (original entire set, training set, testing set), we can see that they have very similar median value of each variables, and the ratios of gamma (signal) to hadron (background) are respectively 1.844, 1.862, and 1.832 which are also very close. Therefore, both of the training and testing data sets are highly representative of the distribution of the entire dataset.

**3. Logistic Regression**



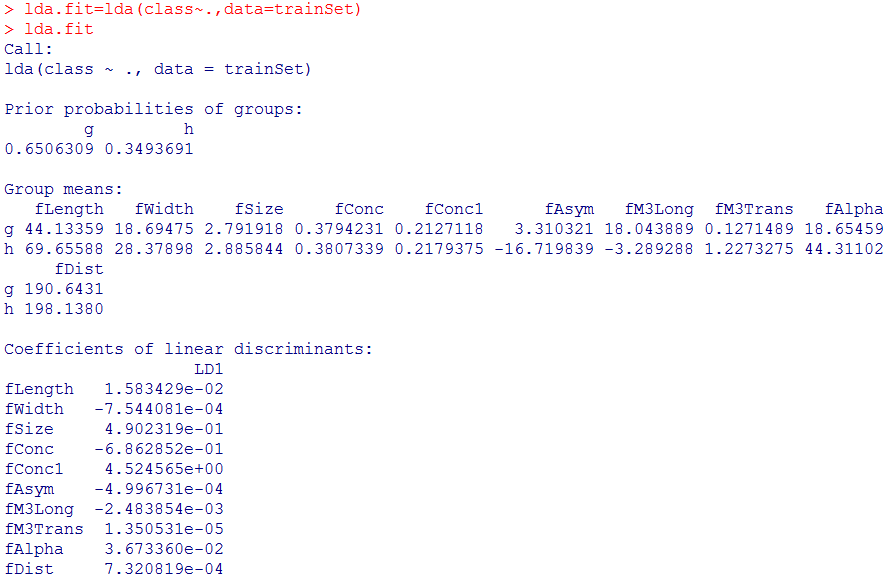


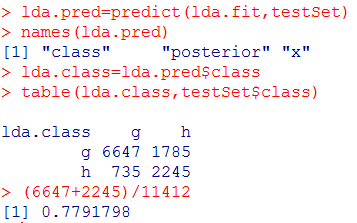
The **testing error rate** in logical regression is **0.212**. It is worth noting that the variables that include “fConc”, “fAsym”, “fM3Trans” and “fDist” have relatively large p-value. If they are removed from logical regression, how about the prediction?



The testing error rate is still **0.212**, which proves that variables of “fConc”, “fAsym”, “fM3Trans” and “fDist” have the negligible function in prediction of the rays class.

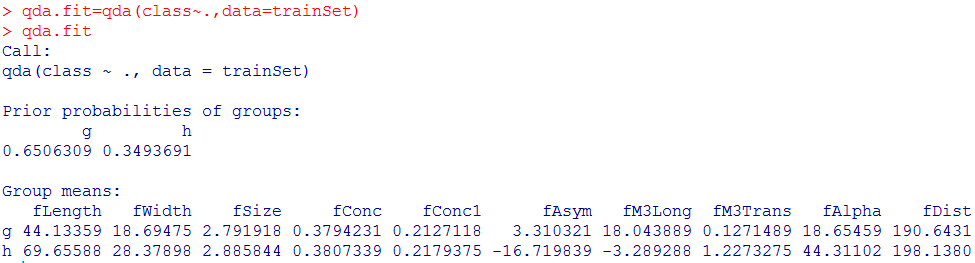
**4. LDA & QDA**

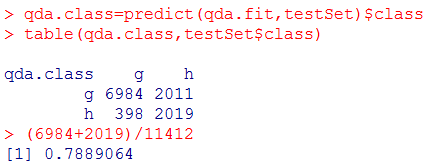
**Linear Discriminant Analysis**



The testing error rate in LDA is 0.221.

**Quadratic Discriminant Analysis**





The testing error rate in QDA is 0.211, which is a little smaller than that in LDA.