**Part 1.1 - Feature Engineering with Feature Subsets (10 points)**

• 1.1.1 Which model had the best RMSE on the training data? (1 point)

* **Ans: artist\_reviewauthor\_releaseyear\_recordlabel\_genre\_danceability\_energy\_key\_loudness\_speechiness\_acousticness\_instrumentalness\_liveness\_valence\_tempo**

• 1.1.2 Which model had the best RMSE on the test data? (1 point)

* **Ans: artist\_reviewauthor\_releaseyear\_recordlabel\_genre\_danceability\_energy\_key\_loudness\_speechiness\_acousticness\_instrumentalness\_liveness\_valence\_tempo**

• 1.1.3 Which feature do you believe was the most important one? Why? (Note: There is more than one perfectly acceptable way to answer this question) (2 points)

* **Ans: reviewauthor** as it is the feature that produced the lowest RMSE on its own.

• 1.1.4 What can we say about the utility of the Spotify features based on these results? (1 point)

* **Ans:** The features are all individually useful and also useful in combination. Each one on its produced a ‘reasonable’ RMSE and combined produced the lowest RMSE.

**Part 1.2 - Feature Engineering with the LASSO (15 points)**

• 1.2.1 - How many new features are introduced by Step 2 above? Provide both the number and an explanation of how you got to this number. (2 points)

* **Ans: 680**

• 1.2.2 - What was the best alpha value according to your cross-validation results? (5 points)

* **Ans: 3.87224703978818e-05**

• 1.2.3 - What was the average RMSE of the model with this alpha value on the k-fold cross validation on the training data? (3 points)

* + **Ans: 0.24625592947243133**

• 1.2.4 - What was the RMSE of the model with this alpha value on the k-fold cross validation on the test data? (5 points)

* + **Ans: 0.2662673832867446**

**Part 1.3 - Interpreting Model Coefficents (15 points)**

* 1.3.1 - How many non-zero coefficients are in this final model? (5 points)
  + **Ans: 468 (print('number of non-zero elements: ',np.nonzero(model.coef\_)[0].size))**
* 1.3.2 - What percentage of the coefficients are non-zero in this final model? (1 point)
  + **Ans: 67.72793053545585 (print('percentage of non-zero elements: ',np.nonzero(model.coef\_)[0].size/model.coef\_.size\*100))**
* 1.3.3 - Who were the three most critical review authors, as estimated by the model? How do you know? (3 points)
  + **Ans: Joe Tangari, Mark Richardson, Stephen M. Deusner (print((pd.concat([training\_data['reviewauthor'],pd.DataFrame(np.power(2,y\_pred\_test)-1,columns=['score'])], axis=1).groupby('reviewauthor').sum()).sort\_values(by=['score'],ascending=False)[2:5]))I**
* 1.3.4 - Who were the three artists that reviewers tended to like the most? How do you know? (3 points)
  + **Ans: Xiu Xiu, Mount Eerie, Guided by Voices (print((pd.concat([training\_data['artist'],pd.DataFrame(np.power(2,y\_pred\_test)-1,columns=['score'])], axis=1).groupby('artist').sum()).sort\_values(by=['score'],ascending=False)[2:5]))**
* 1.3.5 - What genre did Pitchfork reviewers tend to like the most? Which genre did they like the least? (3 points)
  + **Ans:**

**Part 1.4 - “Manual” Cross-Validation + Holdout for Model Selection and Evaluation (25 points)**

1.4.1 Report, for each model, the hyper parameter setting that resulted in the best performance (3 points)

* **Ans:**

1.4.2 Which model performed the best overall on the cross-validation? (3 points)

* **Ans:**

1.4.3 Which model performed the best overall on the final test set? (3 points)

* **Ans:**

1.4.4 With respect to your answer for 1.4.3, why do you think that might be? (Note: there is more than one correct way to answer this question) (1 point)

* **Ans:**

1.4.5 Which model/hyperparameter setting had the highest standard deviation across the different folds of the cross validation? (3 points)

* **Ans:**

1.4.6 With respect to your answer for 1.4.6, why do you think that might be? (Note: there is more than one correct way to answer this question) (2 points)

* **Ans:**

**Part 2.1 - Logistic Regression with Gradient Descent (25 points)**

**2.1.1** - How did you go about selecting a good step size, i.e. one that was not too big or too small? (Note: There is more than one correct answer to this) (2 points)

* **Ans:**

**2.1.2** - What is the condition under which we assume that the gradient descent algorithm has converged in the code here? (2 points)

* **Ans:**

**2.1.3** - What is a different convergence metric we could have used? (Note: There is more than one correct answer to this) (1 points)

* **Ans:**