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

* The feature set ['artist', 'reviewauthor', 'releaseyear', 'recordlabel', 'genre', 'danceability', 'energy', 'key', 'loudness', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo'] had the best training rmse with a value of 1.247.

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

* The feature set ['artist', 'reviewauthor', 'releaseyear', 'recordlabel', 'genre', 'danceability', 'energy', 'key', 'loudness', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo'] had the best test rmse with a value of 1.382.

• 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)

* A single feature alone is not a good way to determine the score in our opinion. But. Still if we have to select a single feature then it will be the review author as out of all the given features in feature set the review author feature had both the least train and test rmse

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

* The spotify features do a great job in differentiating the music type and thus make our job much easier in terms of doing the regression part.

• 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)

* The total number of features introduced after oneHotencoding is 680. We have scaled our continuous variable using standardscalar which does not change the number of rows. The total number of rows which we have is 691 after both oneHotencoding and standardscalar. We have 11 continuous variables and subtracting that from 691 will give us the number of new features added which is 680.

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

* The best value of alpha according to our cross-validation results is 0.0001685.

• 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)

* The average RMSE of the model with this alpha value on k-fold cross validation is 1.122921.

• 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)

* The RMSE of the model with this alpha value on k-fold cross validation on the test set is 1.164979.

• 1.3.1 - How many non-zero coefficients are in this final model? (5 points)

* There are 484 non-zero coefficients in the final model.

• 1.3.2 - What percentage of the coefficients are non-zero in this final model? (1 point)

* Total coefficients in the final scaled model = 691

Total non-zero coefficients = 484

Percentage of non-zero coefficients = 70.04%

• 1.3.3 - Who were the three most critical review authors, as estimated by the model? How do you know? (3 points)

* Matt Lemay, Eric Carr , Alison fields are the 3 most critical review authors. Sorting the scores and calculating the average of each unique author. Selecting the 3 ones with least mean will give us the following answer.

• 1.3.4 - Who were the three artists that reviewers tended to like the most? How do you know? (3 points)

• 1.3.5 - What genre did Pitchfork reviewers tend to like the most? Which genre did they like the least? (3 points)

* They tend to like Jazz the most. They tend to like Pop/R&B the least.

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

* Ridge = (Alpha=10), Knn=(neighbors = 15), DecisionTree=(max\_depth=5)

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

* Ridge regression performed the best with least error.

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

* Ridge regression performed the best with least error.

• 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)

* The mean rmse for ridge regression is the least for the following dataset.

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

* Ridge regression has highest standard deviation across different folds.

• 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)

* It is because of the way ridge regression works. The features are highly dependent on each other and due to multicollinearity it may get diverse values.

• 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)

* The step size for gradient descent should ideally be between 0 and 1 such that it tends towards getting the best value for weights. If we keep the step size small then the loop runs infinitely and if we keep it big then we may not get the best values. Thus, we selected step size as 0.01.

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

* When the square of difference between new and old weights is very small.

• 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)

* We could have ran the loop n(Let’s say 500) times instead of that condition because after 50 or so iterations the value changes at a very slow rate and if not the best we could have gotten values which are very close to the best value.