

Predicting Musical Eras of Songs

Yandi Li, Jiaqi Xue, Yueyao Zhu

Target Application

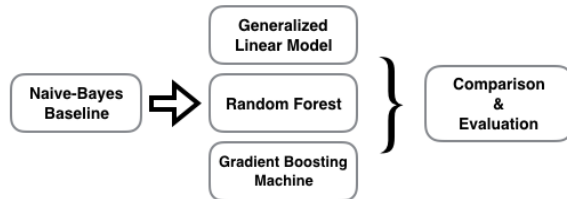
Different musical eras is an important music feature. Certain music has very obvious mark of their time.

The goal is to predict the year in which a certain piece of music was created. So that we can give out the list of music with characteristics belonging to that period no matter it was created at that time or not.

Dataset

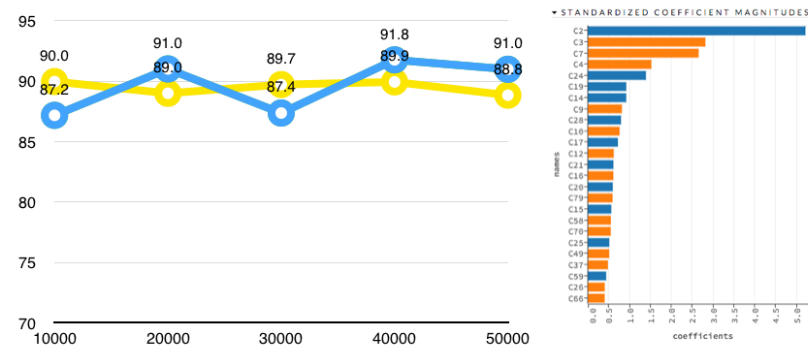
- Dataset used is the Million Song
- Subset for the dataset split into train and test sets
- Dataset contains 90 attributes, "year" ranging from 1922 to 2011
- Features including timbre average and timbre covariance values

Project Workflow



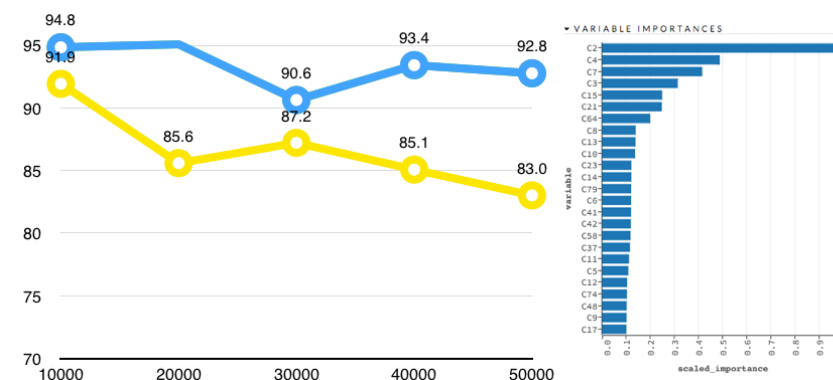
Generalized Linear Model

Generalized linear model is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution.



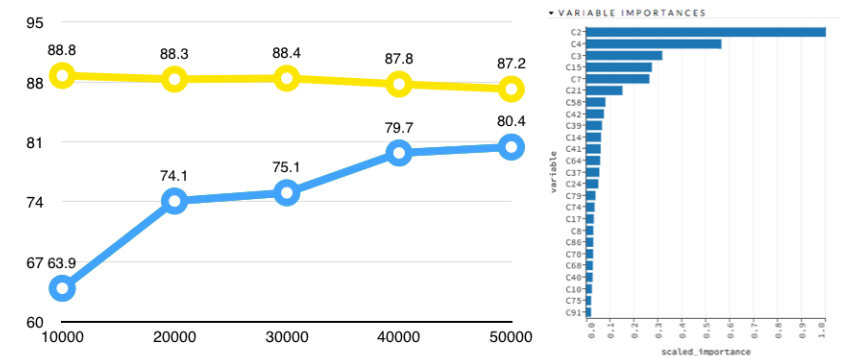
Random Forest

Random forests uses an ensemble of decision trees to make regression and classification models.



Gradient Boost Model

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models.



Conclusion

- For each model, certain variables carry large weights on the resulted prediction (i.e. C2~C7, segment timbre average)
- On the scale of training data tested(50k samples), predictions from three models have similar results

Acknowledgement

Professor Andrew Ng, CS 229 Teaching Staff, LabROSA in Columbia University