Music Rating Prediction

Brian Doolittle and Pratap Luitel

Problem:

How much will a listener like a new song given their...

- User Profile
- Artist Ratings
- Words
- Training Data (example ratings)

Completed for Milestone:

- Preprocessing
- Baseline Models
- Data Analysis

Preprocessing

Data types:
- Binary {0,1}   *i.e.* Gender
- Classes: {0,1,...,n}   *i.e.* Region
- Integer: [0,100]   *i.e.* Rating

Missing Data:
- Empty fields are filled with a marker of -1

Data Overview:
- 50 Artist
- 184 Tracks
- 50928 Users
- 24 times

Baseline Models

Baseline Cross Validation, N = 10

- Used average track rating to predict new track rating
- RMS Error: 21.25

Weighted Average of Features Cross Validation, N = 10

- Predict ratings using weighted average of features
- Learn weights of features
- RMS Error: 18.69

$$\vec{B} = [Artist_{avg}, Track_{avg}, User_{avg}, Time_{avg}]$$
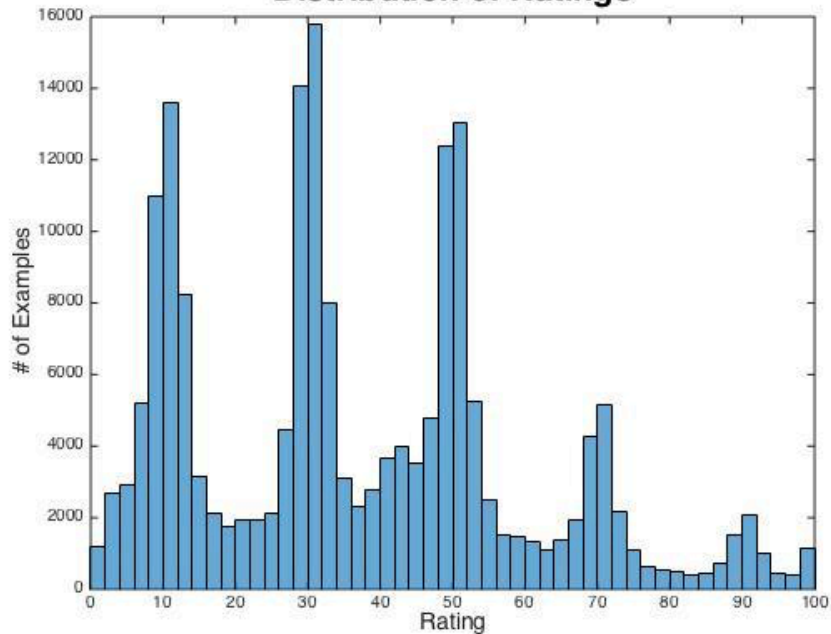
$$\theta_i = \frac{y_i}{B_i}$$

$$\vec{Y}_{pred} = \vec{\theta}\vec{B}$$

Pitfalls:

- Average may not be best feature choice
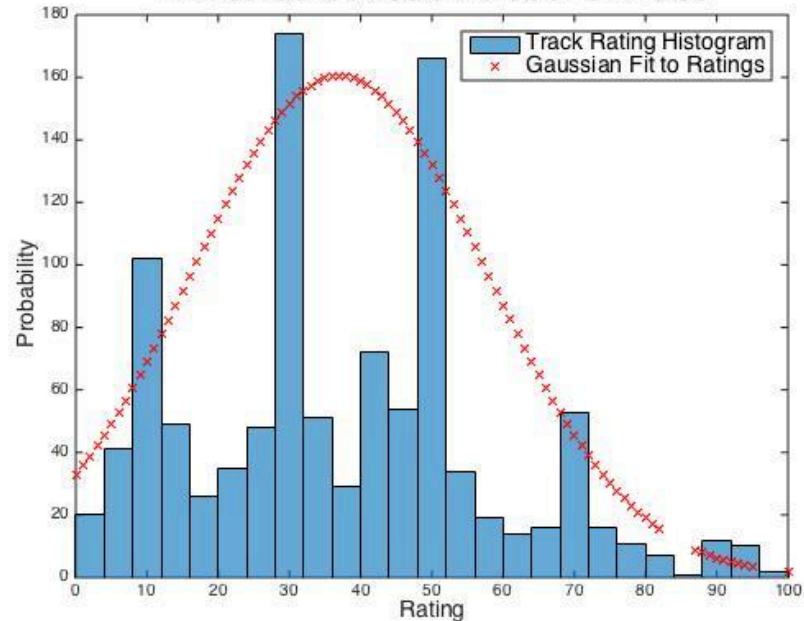- Currently using actual rating to learn weights

Data Analysis

Distribution of Ratings

Average Rating = 36

Truncated-Gaussian Model of Noise

Average Log-Likelihood for Track:
Binomial Distribution = -4722
Gaussian = -4575
Truncated-Gaussian = -4513

Conclusion:

- We will need to model each data type differently
- Baseline Models have poor feature selections
- Ratings may be able to be modeled with classification

Next Steps:

- Incorporate more features in model
- Implement kNN classifier to predict user rating
- Try predicting ratings with classification