Song recommendation with Spotify.

Data Analysis Final Project - Alexis Martin

04/10/2019

Topic

How to create a recommendation algorithm based on 900 Spotify songs!

Spotify has an open API, enabling anyone to access a vast amount of data points from songs.

Questions raised:

- How many features make a good model?
  - 15 chosen
- How to choose features?
  - Mix of Genre classification, Song Technical features & Ratings

Workflow presentation

Get Data

- Spotify API Scraping
- 3 functions used
  - Get playlist tracks
  - Get track features
  - Get artist features

Understand Data

- Merging scraped data:
  - Playlists
  - Track features
  - Artist features
- Data cleaning:
  - Null values
  - Duplicate values
  - Feature engineering
- Data Viz:
  - crosstab scatterplot
  - t-SNE plot

Model Data

- Unsupervised Machine learning:
  - K-means clustering (5)
  - StandardScaler: z=(x-u)/s
  - PCA Analysis
- Recommendation algorithm based on Cosine Similarity
  - scikit-learn

Spotify

Get Data | Understand Data | Model Data

How to create a song dataset?

13 selected Playlists - Genres:

- 70’s
- 80’s
- 90’s
- 00’s
- Pop
- Hip Hop
- Rap
- Rock
- Disco Folk

902 Tracks

Tracks:

- Artist Name
- Artist id
- Song Name
- Song id
- Song Popularity

Track features:

- Danceability
- Energy
- Loudness
- Speechiness
- Acousticness
- Instrumentalness
- Liveness
- Valence

Artist features:

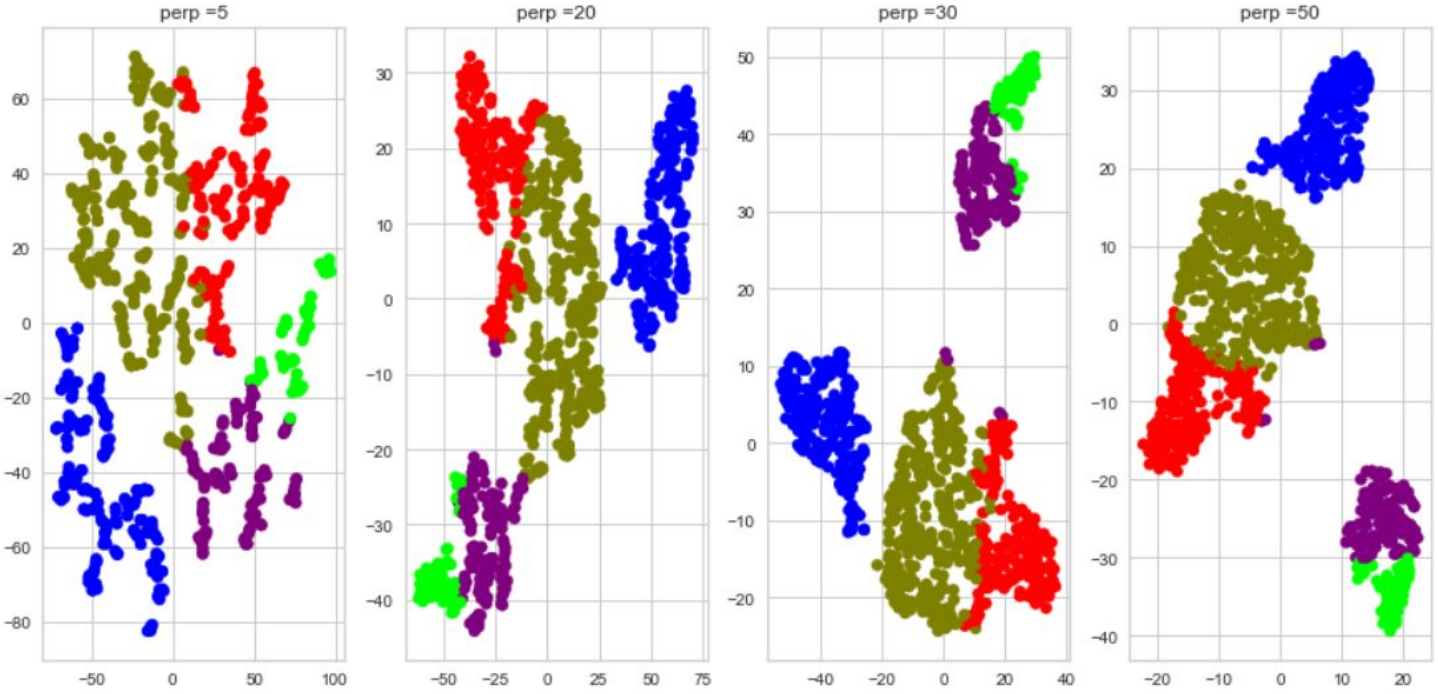
- Artist Popularity
- Genres

Data Visualization for Clustering Problem

t-SNE Plot:

- 5 clusters
- Silhouette score = 0.35
- Clustering influenced by Song and Artist Popularity

=> No clear clustering by genre or song features



Cosine Similarity

Every song is represented as a Vector

- If 2 vectors have an alpha value of 1, they are parallel => a song vector is parallel with itself
- If their value is 0, they are perpendicular => they have nothing in common
- Take one song: 'Hotline Bling'. A song with a score close to 1 is a very similar song
  => Will be recommended

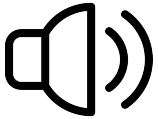
```
0.9996803508846684 Started From the Bottom
0.9706813757551638 Stronger
0.9679248736816763 MIDDLE CHILD
0.9679199364298984 Lollipop
0.9677117803231711 Tha Crossroads
0.9670555360944899 Bop
0.9663262853245249 Dip (feat. Nicki Minaj)
0.9662813095366266 A Milli
0.9655456711086232 Praise The Lord (Da Shine) (feat. Skepta)
0.9645518067440901 Wu-Tang Clan Aint Nuthing ta F' Wit
0.9644775132313446 Fu-Gee-La
0.9643323153494989 La La Land (feat. YG)
0.9642214353365215 The Flute Song
0.9641953181864105 Put It On
0.9640736945489543 Method Man
0.9637727587725821 Ready or Not
0.9629596746050533 Candy Shop
0.9629040913217907 Freaky
0.9627998766656265 Money In The Grave (Drake ft. Rick Ross)
```

Spotify

All this for what???

...

DEMO



Main insights

1. Spotify Track features offer low insight on variability between tracks

2. Popularity measures create a biased model

3. Song recommendations based solely on track features and popularity are not precise. Adding genres of songs makes the model stronger.

Future improvements

1. Create your own new playlist based on the recommendations for every song in an existing playlist

2. Discover new songs from different:
    - Genres
    - Artists
    - Popularity