EE373B Project Report
Can we predict general public's response by studying published sales data? – A Statistical and adaptive approach

Song Hui Chon
Stanford University

Everyone has different musical taste, and a person's preference may change over time, which is often unpredictable. However, general public as a collective unit of individuals may reveal an entirely different pattern of musical taste, and furthermore they may be even predictable. The goal of this paper is to find these patterns which may help us understand how general public, who can be regarded as a group of "average" people in a statistical sense, will respond to new stimuli.

There are three questions this paper has in mind. One is to find any statistically meaningful patterns within data. I used the word "any" here, since I did not know what really to expect. The next question is if we can predict how long an album will stay in chart, given the first few weeks' sales data, using statistical patterns found from the first question. The last question, which was the ultimate motivation of this project, is to see if a new album's position in chart can be predicted on a certain week in the future (such as the 5th week or 12th week), with the first few weeks' sales data. For this, LMS (least mean square) algorithm, a well known adaptive algorithm, has been used.

This paper considers published bi-weekly sales data from the Billboard magazine. Furthermore, I only concentrated on the Top Jazz Albums chart. The results show some interesting correlations, one of which emphasizes the role of marketing. According to my findings, it is probably worth a good investment on marketing before starting sales of an album, since the result shows that the higher the starting position of an album is, the longer it is likely to stay in chart.

INTRODUCTION

Over the years the music industry has seen a growing number of artists and recording companies, as well as of attempts to figure out the secret recipe of a possible hit song. Numerous analyses have been conducted from temporal, acoustical and lyrical perspectives, some of which concentrate on musical similarity and classification, as in [1][2][3] and [4]. There is even a website that claims to have all the formula for musical success [5], though I have not encountered a research paper on this matter from a statistical point of view. The motivation of this project is to detect any statistical patterns in general public's taste of music, by using Billboard Charts data and furthermore to use them to predict an album's success.

In this paper, I used over 3 years of Billboard Top Jazz Albums charts [6]. The objectives here are to see if there is a statistically generic lifecycle model for an album in the genre of Jazz and if there is a correlation between different parameters, such as an album's starting position in the chart with its lifespan. Also, I tried to predict the future, such as how long an album's lifespan will be and what will be the position of an album on a certain week in future.

Before proceeding further, I would like to define two terms – lifecycle and lifespan – which will be used throughout this paper. A lifecycle of an album is a trajectory of the album's weekly positions from the very first week to the very last week in Top Jazz Albums chart, within the time period that was considered for this project. If the album happens to be off-chart for a number of weeks before coming back in to the chart, those off-chart weeks are also considered to be a part of the lifecycle. A lifespan is defined to be how long the lifecycle of an album is, in terms of the number of weeks. Again, if an album happens to be off-chart during one or more times within its lifecycle, the lifespan also includes those weeks off-chart.

EXPERIMENTS

Stanford University has scanned records of Billboard Charts available online from 2002, and I picked Top Jazz Albums chart (a weekly list of No.1 through No.25 in terms of sales rank) for this project. An example of Top Jazz Albums chart is shown in Figure 1. The reason for concentrating on one genre was that I believe it would yield cleaner results that could provide a better insight. Also, there is the reason that Jazz is a genre with unique characteristics, such as that it has a specific audience with a rather well-defined taste and that Jazz audience is a knowledgeable group, in comparison with other more popular genres, like Pop or R&B.

TOP JAZZ

THIS WEEK	LAST WEEK	WEEKS ON CHT	ARTIST / TITLE IMPRINT & NUMBER / DISTRIBUTING LABEL	CERT
1	1	20	#1 19 WKS MICHAEL BUBLE / IT'S TIME 143/REPRISE 48946/WARNER BROS.	●
2	2	3	PAUL ANKA / ROCK SWINGS VERVE 004751/VG	
3	3	41	MADELEINE PEYROUX / CARELESS LOVE ROUNDER 613192	
4	4	39	CHRIS BOTTI / WHEN I FALL IN LOVE COLUMBIA 92872/SONY MUSIC	●
5	5	3	JOHN SCOFIELD / THAT'S WHAT I SAY: JOHN SCOFIELD PLAYS THE MUSIC OF RAY CHARLES VERVE 004360/VG	
6	6	2	HARRY CONNICK, JR. / OCCASION MARSALIS 613313/ROUNDER	
7	11	5	VERA LEE / 83 AND STILL PLAYING WITH THE BOYS S.D.E.G. 1954	
8	7	62	DIANA KRALL / THE GIRL IN THE OTHER ROOM VERVE 001826/VG	●
9	8	3	VARIOUS ARTISTS / PLAYBOY JAZZ: AFTER DARK II CONCORD JAZZ 2751/CONCORD	
10	17	57	RENEE OLSTEAD / RENEE OLSTEAD 143/REPRISE 48704/WARNER BROS.	
11	12	5	JOSHUA REDMAN ELASTIC BAND / MOMENTUM NONESUCH 79864/WARNER BROS.	
12	9	11	TORD GUSTAVSEN TRIO / THE GROUND ECM 004123/UNIVERSAL CLASSICS GROUP	
13	10	2	WAYNE SHORTER QUARTET / BEYOND THE SOUND BARRIER VERVE 004518/VG	
14	14	73	HARRY CONNICK, JR. / ONLY YOU COLUMBIA 90551/SONY MUSIC	■
15	13	2	EDDIE PALMIERI / LISTEN HERE! CONCORD 2276	
16	20	21	VARIOUS ARTISTS / PUTUMAYO PRESENTS: NEW ORLEANS PUTUMAYO 0232	
17	16	24	DAVID SANBORN / CLOSER VERVE 003095/VG	
18	22	12	VARIOUS ARTISTS / VERVE//UNMIXED3 VERVE 004302/VG	
19	15	3	TERENCE BLANCHARD / FLOW BLUE NOTE 78273	
20	18	42	JANE MONHEIT / TAKING A CHANCE ON LOVE SONY CLASSICAL 92495/SONY MUSIC	
21	24	8	KEITH JARRETT / RADIANCE ECM 004314/UNIVERSAL CLASSICS GROUP	
22	RE-ENTRY		ARTURO SANDOVAL / LIVE AT THE BLUE NOTE HALF NOTE 4522	
23	RE-ENTRY		ELDAR DJANGIROV / ELDAR SONY CLASSICAL 92593/SONY MUSIC	
24	19	40	VARIOUS ARTISTS / 20 BEST OF JAZZ MADACY SPECIAL PRODUCTS 5328/MADACY	
25	RE-ENTRY		BOBBIE EAKES / SOMETHING BEAUTIFUL BCI 40960	

FIGURE 1. AN EXAMPLE OF BILLBOARD TOP JAZZ ALBUMS CHART

For this project, I used albums on Billboard Top Jazz Albums charts from August 31, 2002 to January 07, 2006. More specifically, I only considered 293 albums which started and ended the lifecycle during the specific period. Of course, an album whose lifecycle seemed to have ended before January 07, 2006 can still come back to the chart, hence continuing its lifecycle. But there had to be a limit in the data set, especially since this is a first attempt to find patterns in this data. Perhaps I can expand this period of consideration in the future.

Statistical Analysis

The 293 albums showed lifespan of minimum 1 week to maximum 104 weeks. The average was 15.5 weeks and median 9 weeks. The histogram of lifespan is shown in FIGURE 2. Out of 293 albums, almost 40 albums had lifespan of 1-2 weeks and 5 albums had lifespan of over 99 weeks. Since it seemed logically appropriate to group data together by the length of the lifecycle (or lifespan), I grouped the albums into 13 different categories, which are 1-2, 3-5, 6-9, 10-14, 15-19, 20-24, 25-29, 30-35, 40-49, 50-53, 60-69, 88-90 and 99+ weeks. There are some intervals which were not included in the categories (such as 54-59 weeks) and it is because there was no single data that fell into the category.

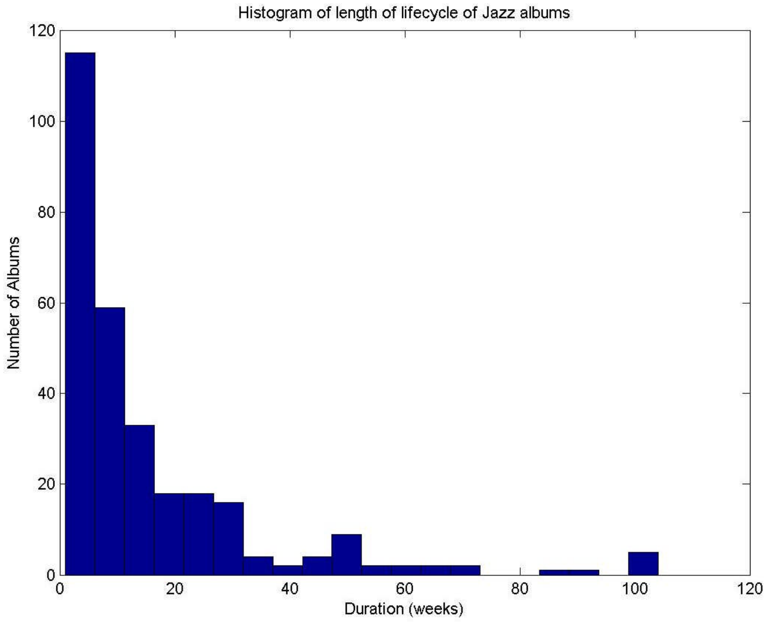

Figure 2. Histogram of lifespan of Jazz albums

For statistical analysis, Microsoft Excel and Mathworks Matlab programs were used. Before the experiments, the hypothesis was that we would see something close to a Gaussian curve for a generic lifecycle of an album, starting at a low position in the chart, climbing to higher places before dropping down to off chart. However, the results show that many albums exhibit a similar trend of starting near its peak position and gradually climbing down. There is a strong correlation between the starting position of an album in the chart and the duration of its lifecycle. For example, as can be seen in Figures 3 and 4, more than a half of albums that had only 1 week lifespan started at below 20, while 3 out of the 5 albums that had over 99 weeks of lifespan started at position 1.

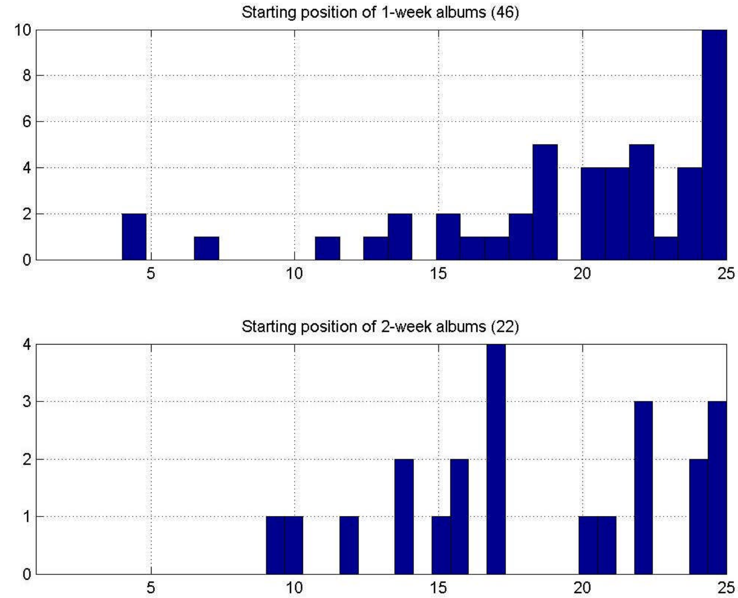

Figure 3. Starting positions of 1- and 2-week lifespan albums

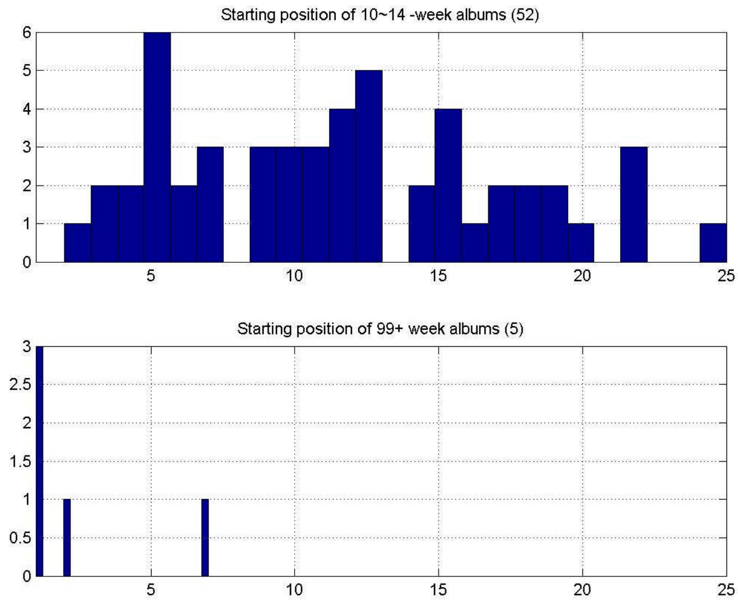

Figure 4. Starting positions of 10~14- and 99+-week lifespan albums

After the 293 albums were categorized into thirteen groups according to their lifespan, each group's average lifecycle was obtained. Figure 5 illustrates average lifecycles of 12 groups (excluding the group of 1-2 week(s)). Note that the starting points of lifecycle decrease as the lifespan increase. It is also noticeable that nearly all the lifecycles exhibit a linear descent and that even when there is a deviation from the trend of linear decay (i.e. a temporary upward move in the chart position), it is never as high as the original peak position that was achieved during the first few weeks of lifecycle. The only exception we can see is the case for 88~92-week lifespan albums, which seems to show a fluctuating tendency, neither increasing nor decaying on average. This comes from the fact that a limited amount of data was considered for this project, that there were only two albums that fell into this category. I believe that this category will also show a linear decaying trend as in other cases, had more data been collected. The Matlab code for statistical analysis is shown in the Appendix with minimum distance algorithm (jazz.m).

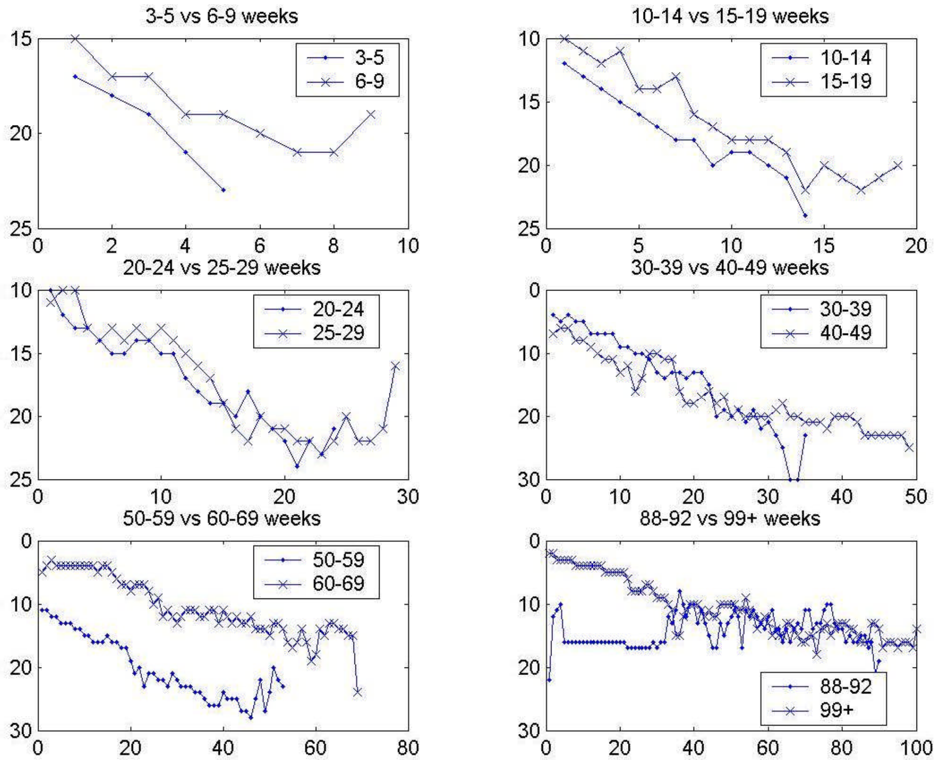

FIGURE 5. AVERAGE LIFECYCLE GROUPED BY LIFESPAN

Minimum Distance Algorithm and Lifespan Prediction

One of the questions this paper answers is whether we can predict the lifespan of a new album, given its first few weeks of sales data. For this, a minimum distance algorithm has been used. This algorithm calculates Euclidean distances between the first few weeks' sales history of a new album considered and the same number of weeks from the average lifecycle patterns from the thirteen categories and determines the expected number of lifespan with the minimum distance.

For example, consider an album whose lifecycle has been [9 15 13 11] and we are interested in how long it will be in chart. Since the given data vector is only 4-dimensional, this cannot be compared with other "average" vectors, whose dimensions are greater than 4. The average vector in 1~2-week lifespan category will be excluded from this calculation since the album of question is already on its 4th week into its lifecycle.

Each average vector's dimension was reduced to 4 by taking the first four numbers (or positions) in each of the average vectors. Then the Euclidean distance from each of these 4-dimensional vectors to the new vector in question was calculated, and the minimum distance obtained, along with the index of the "average vector" which corresponds to it. The index is an estimation of how long this new album will be in chart. Going back to our example, the index of the minimum distance turns out to the average vector of the 20~24-week category. Therefore, this album is expected to stay in chart for 20~24 weeks.

After a preliminary statistical studies using Microsoft Excel, the results were stored in text files and Matlab program (jazz.m) was used for actual calculation. The Matlab code is shown in Appendix.

Table 1. Four albums considered for lifespan estimation and their first 10-week lifecycles

Artist	**Album**	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**
Chris Botti	To love again	1	1	2	3	3	4	4	4	3	5
Jane Monheit	Season	9	15	13	11	10	11	8	11	12	15
Shirley Horn	But Beautiful	17	8	15	16						
Various Artists	Martha Stewart Living Music: Jazz for the holidays	18	9	10	8	5	6	6	7	7	10

Table 1 lists four recent albums, considered for this lifespan estimation. These albums were not among the 293 albums which built the average lifecycle vectors. For the album #1 (*To love again* by Chris Botti), the algorithm predicted it to perform very strong (to 99+ weeks). It is still doing very well on the chart, at the number 3 spot on the 30th week, as of the last week of May 2006. The album #2 (*Season* by Jane Monheit) showed a very strong performance of 12 weeks (10th position on week 12), so the algorithm expected it to last for over 99 weeks, where it went off chart after the 12th week. The album #3 (*But Beautiful* by Shirley Horn) was expected to last 10-14 weeks on chart, while in reality it has not come back to chart after the 4th week (as of the last week of May 2006). For the album #4 (*Martha Stewart Living Music: Jazz for the Holidays* by Various Artists), it stayed pretty high in chart for the first 12 weeks, so the algorithm estimated that it would stay in chart for at least 40 weeks. This album went off-chart after the week 12.

Albums #2 and #4 would not be a normal case, but it makes sense seeing that they both were Christmas albums and their lifecycles started in November. After January, we expect to see almost no Christmas albums on the chart.

Overall, the algorithm seems a bit too optimistic. This probably comes from the fact the model is very simple (almost too simple) and it probably needs to consider other factors such as the starting date (when the album first came on chart) to put the seasonal factor into consideration. The seasonal factor (e.g. for Christmas or for Valentine's day) would affect an album's lifecycle significantly, independent of how strong it performs while on chart.

Adaptive Algorithm and Lifecycle Prediction

The ultimate question in mind for this paper was whether a particular album's performance can be predicted for the next week, given the first few weeks of sales data. Least Mean Square (LMS) algorithm [7], a well-known adaptive algorithm, was used for this. An adaptive model with 104 adjustable weights (the longest lifespan amongst the 293 albums) was trained with the data of 293 albums, with a specific week in mind (for example, week 5 or week 30). Again, the Matlab code is in Appendix (jazz_lms.m). The results are slightly different with the specific week considered, but the model predicted that by week 30 most albums will be off chart. The four albums on Table 1 were considered also for this experiment. For the album #1, which is still very strong in chart, the model expected it to be at 18th by week 5 and off chart by week 30, which is quite different from real data. And for the other three albums, the model predicted them to be at 21st by week 5 and off chart by week 30. The system's prediction was a bit closer to what's observed in the real sales data in these cases.

The discrepancy between what's expected from the system and what's observed in real sales data probably comes from the fact that the model which was built for the experiment was not complex enough to handle the overall complexity of the real data. Also, there is another factor in data that all the off-chart positions were assumed to be 30 since there is no access to the real sales data when an album goes off-chart. I presume that a better estimation would have been possible with some kind of interpolation technique used for off-chart positions, but it was not possible with the limited time for this project. The next step will be to refine this model for a better prediction, possibly using a different adaptive technique and to extend the scope of this project to other charts and see if similar patterns can be found.

ANALYSIS OF RESULTS

A statistical and adaptive analysis has been performed to find patterns in Billboard chart data as a measure of general public's response over time. 40 months' data from Billboard Top Jazz Albums chart were used for the project. This project focused on one genre, since it would help find cleaner and more coherent patterns which may give better insights.

There were a number of assumptions made on the data, including that all the off-chart position was set to 30. This was a very crude way of "filling the gaps" in data, because the exact off-chart positions for each album considered was not available from the data considered.

Interesting patterns were observed as a result of statistical analysis. The 293 albums considered were categorized into 13 different groups, according to their lifespan. Then an average lifecycle was calculated for each category. As noticed earlier, there is a strong correlation between the starting position of an album and how long it stays in chart. Another interesting finding was that most groups showed a consistently linear decay in lifecycle, after debuting at their near-peak positions.

Using the statistical analysis result, a new album's lifespan and lifecycle were estimated. To calculate how long an album is likely to stay in chart given a few weeks of sales data, a minimum distance algorithm was used. LMS (least mean square) algorithm, a famous adaptive algorithm, was used to predict the next week's position of an album. Both estimates turned out to be a bit too optimistic, though there were some cases where the estimates were closer to real data than others.

While I was going through the data, I also noticed that the albums of "famous" artists did very well. For example, Diana Krall is a famous Jazz singer and the albums would start lifecycles at number 1, staying in chart for many weeks. Another example is Elvis Costello, who is well-known though not as a Jazz artist. He had an album that started its lifecycle at number 1 and remained in chart for 28 weeks. Peter Cincotti, whose first album received raves from critics, still did pretty well with his second album, even though the second one was regarded as a disappointment. These examples show the power of marketing and publicity – people will be more eager to buy records of artists whose names they heard of.

Another observation (but with no proof) is that however great an artist is regarded, he/she cannot compete well when they are already dead. Among the 293 albums I analyzed, there were quite a number of albums by big Jazz names such as Ella Fitzgerald, Louis Armstrong and Miles Davis. Those albums could have been of "a better quality" musically, but still lost the battle in popularity. My conjecture for this is that the general public want to be in touch with "what's happening now" instead of being educated by great masters.

CONCLUSION AND FUTURE WORKS

Some interesting patterns were found from the experiments. After categorizing the 293 albums into thirteen categories, an average lifecycle was calculated for each category. This became the basis for lifespan and lifecycle predictions, which produced mixed results.

The "average lifecycle" for each category showed a very similar characteristic of linear decay from the onset, in contrast to my initial hypothesis that the average lifecycle would be something close to a Gaussian bell curve. Also, a strong correlation was found between the starting position of a lifecycle and its lifespan. For example, albums with very short lifespan tend to debut at below 20, while long-lasting albums tend to debut at above 5. From these results, it can be said that marketing before the start of sales of an album is quite important, since these patterns indicate that the higher the starting position is, the longer it will stay in chart.

Using minimum distance algorithm and LMS algorithm, a prediction was attempted on a new album's lifecycle and lifespan. With a limited set of data considered, the predictions were of mixed results, some very different from the real data and some others closer. There are various other adaptive techniques available for analysis. In the future, I would like to try them for a better prediction.

This project produced some very interesting result even though it was done with a very limited set of data and engineering models which are almost too simple. I believe that if I had used a more realistic method to "fill in the gap" in data such as quadratic or cubic interpolation, the estimation result would have improved significantly. That would be something I would like to change in the future. Also, I may want to categorize the data into more groups, therefore being able to analyze more detailed patterns in each group. Perhaps it will yield a result that can be fitted with a line.

I believe that this analysis showed a unique characteristic from the fact that the considered genre was Jazz. Certainly other genres will have other unique characteristics, which should be left to future work. I don't expect myself to repeat this analysis on every chart in Billboard magazine, but I would like to consider a bigger set of data, from possibly mixed genres.

Unlike other genres such as Country, Jazz is a genre with both vocal and instrumental (or non-vocal) music. I would like to further study to see whether there are different patterns in vocal music from instrumental music within the genre of Jazz.

Also, I would like to consider some factors which were not considered for this project, such as skip rate (how often an album goes off chart) and start date (the first day when the album enters the chart). These may help improve my models better handle season-specific albums, for Christmas or Valentine's Day, for example. With these improvements, I hope I will encounter many more interesting patterns.

REFERENCES

[1] P. Cano, M. Koppenberger, and N. Wack, Content-based audio recommendation, ACM Multimedia 2005

[2] A. Berenzweig, B. Logan, D.P.W. Ellis, and B. Whitman, A large-scale evaluation of acoustic and subjective music similarity measures, In *Proceedings International Conference on Music Information Retrieval (ISMIR)*, pp 103-109, 2003b

[3] J.T. Foote, Content-based retrieval of music and audio, In *SPIE*, pp 138-147, 1997

[4] R. Dhanaraj and B. Logan, Automatic Prediction of Hit Songs, In *Proceedings International Conference on Music Information Retrieval (ISMIR)*, 2004

[5] Hit Song Science on http://www.polyphonichmi.com/technology.html

[6] Billboard Magazine, Top Jazz Albums chart, from 08/31/2002 to 01/07/2006.

[7] Adaptive Signal Processing, by Widrow and Sterns, Prentice Hall, 1985

APPENDIX

Two Matlab scripts used for this project are shown below. The file jazz.m is for statistical analysis and for minimum distance algorithm. The second file jazz_lms.m implements LMS algorithm with 104 weights to predict an album's position on a specific week. Both files were executed on Matlab version 6.5.

```
% jazz.m
%
% For EE373B Project
% Spring 2006
% SongHui Chon

% This matlab code performs basic statistical analysis on data
% and minimum distance algorithm to preduct a new album's lifespan.

% For Question1: Statistical analysis of data
clear all; close all;
weeks = load('jazzDuration.txt'); % a file with lifespan of 293 albums
mean(weeks);            % = 15.5427
median(weeks);          % = 9

figure;
hist(weeks, 20); xlabel('Duration (weeks)'); ylabel('Number of Albums');
title('Histogram of length of lifecycle of Jazz albums');

% Starting position of albums according to the length of lifecycle
wk1pos = load('jazz1wks.txt');      % 1 week duration albums' starting position
wk1pos = wk1pos';
mean(wk1pos);           % = 19.6522
median(wk1pos);         % = 21

wk2pos = load('jazz2wks.txt');      % 2 week duration albums' starting position
mean(wk2pos(:, 1));         % = 18.3636
median(wk2pos(:, 1));       % = 17
```

```
wk10pos = load('jazz10_14wks.txt');% 10-14 week duration albums' starting position
wk10pos = wk10pos';
mean(wk10pos);                  % = 11.6346
median(wk10pos);                % = 12

wk99pos = load('jazz99wks.txt');    % 99+ week duration albums' starting position
wk99pos = wk99pos';
mean(wk99pos);                  % = 2.4000
median(wk99pos);                % = 1

% Plot starting position vs. lifespan
figure;
subplot(211); hist(wk1pos, 25); grid on; axis([1 25 ylim]);
title('Starting position of 1-week albums (46)');
subplot(212); hist(wk2pos(:, 1), 25); grid on; axis([1 25 ylim]);
title('Starting position of 2-week albums (22)');

figure;
subplot(211); hist(wk10pos, 25); grid on; axis([1 25 ylim]);
title('Starting position of 10~14 -week albums (52)');
subplot(212); hist(wk99pos, 25); grid on; axis([1 25 ylim]);
title('Starting position of 99+ week albums (5)');

% Get average and median data
temp = load('wks3_5pos.txt');       % albums of 3-5 weeks lifecycle
avg3 = temp(1,:);

temp = load('wks6_9pos.txt');       % albums of 6-9 weeks lifecycle
avg6 = temp(1,:);

temp = load('wks10_14pos.txt');       % albums of 10-14 weeks lifecycle
avg10 = temp(1,:);

temp = load('wks15_19pos.txt');       % albums of 15-19 weeks lifecycle
avg15 = temp(1,:);
```

```
temp = load('wks20_24pos.txt');       % albums of 20-24 weeks lifecycle
avg20 = temp(1,:);

temp = load('wks25_29pos.txt');       % albums of 25-29 weeks lifecycle
avg25 = temp(1,:);

temp = load('wks30_35pos.txt');       % albums of 30-35 weeks lifecycle
avg30 = temp(1,:);

temp = load('wks40_49pos.txt');       % albums of 40-49 weeks lifecycle
avg40 = temp(1,:);

temp = load('wks50_53pos.txt');       % albums of 50-53 weeks lifecycle
avg50 = temp(1,:);

temp = load('wks60_69pos.txt');       % albums of 60-69 weeks lifecycle
avg60 = temp(1,:);

temp = load('wks88_90pos.txt');       % albums of 88-90 weeks lifecycle
avg88 = temp(1,:);

temp = load('wks99pos.txt');          % albums of 99+ weeks lifecycle
avg99 = temp(1,:);

figure;
subplot(321); plot(-avg3, '.-'); hold on; plot(-avg6, 'x-');
title('3-5 vs 6-9 weeks'); legend('3-5', '6-9', 0);
subplot(322); plot(-avg10, '.-'); hold on; plot(-avg15, 'x-');
title('10-14 vs 15-19 weeks'); legend('10-14', '15-19', 0);
subplot(323); plot(-avg20, '.-'); hold on; plot(-avg25, 'x-');
title('20-24 vs 25-29 weeks'); legend('20-24', '25-29', 0);
subplot(324); plot(-avg30, '.-'); hold on; plot(-avg40, 'x-');
title('30-39 vs 40-49 weeks'); legend('30-39', '40-49', 0);
subplot(325); plot(-avg50, '.-'); hold on; plot(-avg60, 'x-');
title('50-59 vs 60-69 weeks'); legend('50-59', '60-69', 0);
subplot(326); plot(-avg88, '.-'); hold on; plot(-avg99, 'x-');
```

```
title('88-92 vs 99+ weeks'); axis([0 100 ylim]); legend('88-92', '99+', 0);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% For Question2: Prediction of lifespan using minimum distance algorithm
% Vector distance calculation with input vectors
in4 = load('input4wks.txt');       % first four weeks of first 4 rows (from sheet3)
in10 = load('input10wks.txt');    % first ten weeks of rows 1, 2 and 4 (from sheet3)

dist4avg = -100*ones(4, 12);       % distance to the avg vectors considering first four
weeks
dist10avg = -100*ones(3, 10);     % distance to the avg vectors considering first ten
weeks

mindist4avgindex = zeros(1, 4); % index of minimum distance in dist4avg vector
mindist10medindex = zeros(1, 3);% index of minimum distance in dist10avg vector

for i=1:4     % calculate distance between 4-week input vectors with avg vectors
      dist4avg(i, 1) = norm(in4(i,:) - avg3(1:4));
      dist4avg(i, 2) = norm(in4(i,:) - avg6(1:4));
      dist4avg(i, 3) = norm(in4(i,:) - avg10(1:4));
      dist4avg(i, 4) = norm(in4(i,:) - avg15(1:4));
      dist4avg(i, 5) = norm(in4(i,:) - avg20(1:4));
      dist4avg(i, 6) = norm(in4(i,:) - avg25(1:4));
      dist4avg(i, 7) = norm(in4(i,:) - avg30(1:4));
      dist4avg(i, 8) = norm(in4(i,:) - avg40(1:4));
      dist4avg(i, 9) = norm(in4(i,:) - avg50(1:4));
      dist4avg(i, 10) = norm(in4(i,:) - avg60(1:4));
      dist4avg(i, 11) = norm(in4(i,:) - avg88(1:4));
      dist4avg(i, 12) = norm(in4(i,:) - avg99(1:4));

end

for i=1:4
      mindist4avgindex(i) = find(dist4avg(i,:) == min(dist4avg(i,:)));
end
% mindist4avgindex = 12 5 3 11
```

```
% (which means 99+ weeks, 20-24 weeks, 10-14 weeks, 88-90 weeks)

for i=1:3     % calculate distance between 10-week input vectors with avg vectors
      dist10avg(i, 1) = norm(in10(i,:) - avg10(1:10));
      dist10avg(i, 2) = norm(in10(i,:) - avg15(1:10));
      dist10avg(i, 3) = norm(in10(i,:) - avg20(1:10));
      dist10avg(i, 4) = norm(in10(i,:) - avg25(1:10));
      dist10avg(i, 5) = norm(in10(i,:) - avg30(1:10));
      dist10avg(i, 6) = norm(in10(i,:) - avg40(1:10));
      dist10avg(i, 7) = norm(in10(i,:) - avg50(1:10));
      dist10avg(i, 8) = norm(in10(i,:) - avg60(1:10));
      dist10avg(i, 9) = norm(in10(i,:) - avg88(1:10));
      dist10avg(i, 10) = norm(in10(i,:) - avg99(1:10));
end

for i=1:3
      mindist10avgindex(i) = find(dist10avg(i,:) == min(dist10avg(i,:)));
end
% mindist10avgindex = 10 7 6 (which means 99+ weeks, 50-59 weeks, 40-49
% weeks)

==================================================================

% jazz_lms.m
%
% For EE373B Project
% Spring 2006
% SongHui Chon

% This matlab code implements LMS algorithm with 104 adaptive weights.

% For Question3: Prediction of position on a specific week

clear all; close all;

X = load('jazzLMS.txt');      % 293 columns of input data
X = X';                        % transpose X
```

```
[L, K] = size(X);     % l=1,...,L (instead of l=0,...,L) and k=1,...,K
                            % L = # of weeks, K = album index
mu = 0.000001;
W_k = zeros(L, 1);
for i=1:L
    d_k = X(5, i);                    % position at 5th week
    X_k = X(:, i);                    % input vector
    y_k = X_k'*W_k;                   % weighted input sum
    e_k = d_k - y_k;                  % error
    W_k = W_k + 2*mu*e_k*X_k;         % weight adjustment
end

in4 = load('input4wks.txt');      % first four weeks of 4 albums considered (from sheet3)
in10 = load('input10wks.txt');    % first ten weeks of 3 albums considered (from sheet3)

in4ext = [in4, 30*ones(4, L-4)];   % extend the vector by appending 30 (off-chart
position)
in10ext = [in10, 30*ones(3, L-10)]; % extend the vector by appending 30 (off-chart
position)

est4 = -1*ones(4, 1);              % position estimation using first four weeks' data
est10 = -1*ones(3, 1);             % position estimation using first ten weeks' data
est4ext = -1*ones(4, 1);
est10ext = -1*ones(3, 1);

for i=1:4
    est4(i) = in4(i,:)*W_k(1:4);
    est4ext(i) = in4ext(i,:)*W_k;
end
for i=1:3
    est10(i) = in10(i,:)*W_k(1:10);
    est10ext(i) = in10ext(i,:)*W_k;
end

round(est4ext')
round(est10ext')
```