Detection of Churned and Retained Users with Machine Learning Methods for Mobile Applications

Merve Gençer[1], Gökhan Bilgin[2], Özgür Zan[1], and Tansel Voyvodaoğlu[1]

[1] Done Info. and Com. Systems Istanbul, Turkey
{merve,ozgur,tansel}@donetr.com
[2] Yildiz Technical University, Dept. of Computer Engineering, Istanbul, Turkey
gbilgin@yildiz.edu.tr

Abstract. This study aims to find the different behavior patterns of churned and retained mobile application users using machine learning approach. The data for this study is gathered from the users of a mobile application (iPhone & Android). As a machine learning classifier Support Vector Machines (SVM) are used for evaluating in the detection of churned and retained users. Several features are extracted from user data to discriminate different user behaviors. Successful results are obtained and user behaviors are classified with 93% and 98% accuracy. From the diversity perspective, results of this study can be used to evaluate the differences of churned and retained users in terms of diverse user groups.

Keywords: Machine learning, SVM, mobile applications, churned and retained users, diversity applications, classification, mobile devices, push notification, user experience.

1 Introduction

Number of mobile applications in each mobile operating system is growing rapidly. For example, in App Store of Apple, there are over 1 million iOS based mobile applications (as of October 2013) [1]; similarly in Google Play, there are over 1 million android based mobile applications too [2]. As an emerging market, there are more than 100,000 apps in the Windows Marketplace [3].In the statistics provided above, due to high number of mobile applications, it is not easy for a developer company to compete in this market. In addition to this difficulty, it is also costly to acquire a new customer. On the other hand, most of the revenues come from loyal users where it is less costly to retain users. Thus, mobile application development companies should find better strategies to keep their customers. For this reason, it is important to monitor customer behaviors and interests via analytics tools.

However, current analytic tools do not provide forecast information. Our main motivation is to go beyond those analytic tools and detect churned and retained users with their usage characteristics for a mobile application. We will focus on the task of separability of these characteristics with machine learning methods. We will use the data in our previous study [4] in which we have collected 1 million user dynamic

A. Marcus (Ed.): DUXU 2014, Part II, LNCS 8518, pp. 234–245, 2014.

information, 5.5 million user action information and 30.000 user static information during 17th November 2012 - 17th July 2013 and we aim to extend our previous work.

If and only if, a manager is able to forecast these churned and retained mobile users, then s/he will be able to plan company's existing product improvement or new product development duly. For this reason, we need an objective evaluation system. In this study, we are going to classify churned and retained users according to their usage characteristics in a pilot mobile application (as an iPhone and Android application) called "Mobil Lig". In this study, a kernel based classifier method, namely Support Vector Machines (SVM), will be utilized. SVM is chosen due to its robustness and the power of separation both linear and nonlinear distributions/features.

From the diversity perspective, results of this study can be used to evaluate the differences of churned and retained users in terms of diverse user groups. For example, if it is detected that churned users mainly compose of a specific group (e.g. elder people, deaf people, color blind people), then this information can be used to improve user experience design considering these diverse user groups.

In literature, customer churn is also known as customer attrition or customer turnover. It basically defines the loss of customers. Generally, customer churn does not depend on only a single reason, usually there are various reasons. Customer churn is a problematic study area because of the large quality of data and the difficulty of modeling data distributions. In [5], non-linearity in customer churn and large data sets issues has been studied to find a systematic way of dealing with these problems. Many researchers have studied customer churn from the point of financial stability. For example in [6], subject of 'predictors of churn and use of customer relationship management (CRM) systems' has been investigated for a solution. The early detection of potential churners for a European pay-TV company has been targeted for increasing retention rates in [7]. An analytical CRM approach is presented by using the pay-TV customers' data in same study. Various data mining techniques have been evaluated to assign a 'propensity-to-churn' score for each subscriber of a telecom operator in [8]. In the study, decision tree and neural network techniques have been utilized to predict models by using the customer demographics, billing information, contract/service status, call detail records, and service change log as features. In another work, improvement of customer attrition prediction is purposed by integrating emotions from client/company interaction emails using several machine learning methods [9]. In the study, logistic regression, support vector machines and random forests classification methods are used to define churners and non-churners classes. Data mining and machine learning employ the same techniques frequently and roughly deals with known and unknown properties of data respectively [10]. In [11], a data mining approach is investigated for retailing bank customer attrition analysis to retain existing customers and reach new prospective customers. In a similar study area, data mining methods are employed to discover the current spending pattern of customers and trends of behavioral change for preventing customer attrition [12]. A Bayesian partition model for customer attrition is performed for bank customers with a large retail banking dataset in [13].

In the proposed study, detection of churned and retained users is aimed with utilizing machine learning methods for mobile applications. A kernel based classifier, namely SVM, is chosen because of its robust mechanism with high dimensional and high volume data. SVM is also a powerful method for revealing nonlinear data structure for churn and retaining patterns. The data for this study is gathered from the users of a mobile application (iPhone & Android). Several features are extracted from user data to discriminate different user behaviors. Successful results are obtained and user behaviors are classified with 93% and 98% accuracy.

2 Methodology

This study aims to find the different behavior patterns of churned and retained mobile application users in order to understand when a mobile application user will quit using the application. When this behavior (continue or quit) is forecasted, this information can be an input for another decision. For example, if the user is detected to continue using the application then her/his next time of usage can be detected, or if the user is detected to quit using the application, then total number of users who may quit are extracted which would help the company foresee the life cycle of the application before it is too late.

In our previous study [4] we had developed a mobile application for both iOS and Android enabled mobile clients and collected over 1 million anonymous user data. Then we pre-processed those data (outlier detection, noise detection, filtering and normalization) for preventing the database from outlier samples. Then we prepared the data for providing appropriate inputs for the machine learning algorithms. The methodology that was used in our previous study is given in Figure 1. In this study, the subsets of these preprocessed data will be used with new features and again machine learning algorithms will be applied. Since our data is already ready, we continued with analysis of these data and explained findings in the next section.

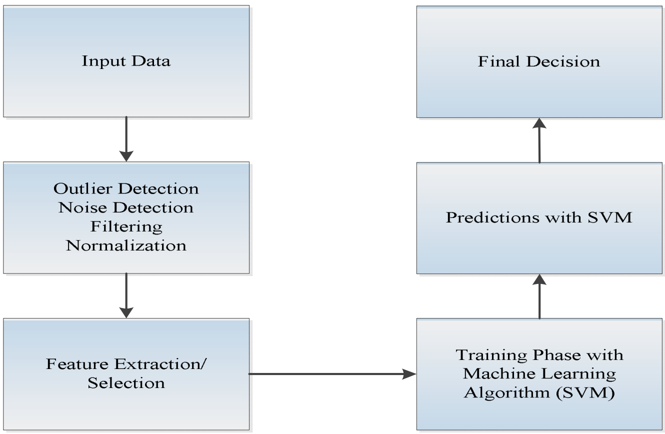

Fig. 1. Flow chart of general machine learning system framework

2.1 Data Sets and Data Collected

Users who are not using the mobile application since 30 days are assumed to be churned users. First, all the users in the database are queried for their start dates and last usage dates, and according to the last usage dates, churned and loyal users are extracted. Thus, mainly two subsets are extracted: churned users who quit the application (i.e. did not use the application more than 30 days), and retained users who have at least one usage in the last 30 days. The data sets used in this study are given as follows:

- Churned Users (E)
  - who have used the application only for one day (E1)
  - who have used the application for more than 7 days (E7)
- Retained Users (N)

The types of data collected are:

User Static Information. It simply covers the general information about the device and the operating system of the user. Device name and model, operating system name and version, application version, user tracking library version are gathered and sent to the database.

User Action Information. It covers the most important features of the database. All the actions of the user in the application are tracked and recorded to the database.

User actions include date and time information of the following actions:

- Openings and closings of the application,
- Entering background and foreground,
- Clicked buttons,
- Opened pages

Names of the buttons and pages of interest are also recorded. User location is also tracked as a user action for every 50 meters change.

User Dynamic Information. It covers the information about the device, that can be changed according to the environment. Battery status (plugged, unplugged, charging, discharging, not-charging), battery level, brightness level, headphone status (plugged, unplugged), internet connection status (Wi-Fi or Cellular), volume level of the application and gyroscope information are gathered as dynamic information from the users.

3 Support Vector Machines

SVM is a robust and widely used kernel based classification algorithm in machine learning society [14]. In kernel based learning, data samples are mapped to high dimensional space by a kernel function to separate data. In this way, inseparable data points in low dimensional feature space can be separated in high dimensional kernel space. SVMs are firstly introduced for the binary classification problems with n dimensional feature vector x_i and binary class label y_i. SVM forms a decision plane between the samples of different classes by optimizing n dimensional hyperplane. Optimization process defines the closest training samples to this hyperplane which are called as support vectors. In case of nonlinear distributions, kernel functions are used. The idea of so called 'kernel trick' is a transformation of data into a higher dimensional space ($\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m, m > n$) where binary classification can be achieved linearly again [15]. Kernel functions in SVM perform an inner product in the higher dimensional space. Support vectors can be found by the solution of the following optimization problem that maximizes (1) with subject to (2).

$$\sum_{u=1}^{N} \alpha_u - \frac{1}{2} \sum_{u=1}^{N} \sum_{v=1}^{N} \alpha_u \alpha_v y_u y_v K(x_u, x_v) \tag{1}$$

Kernel functions $K(x_u, x_v)$ does not require an explicit definition of ϕ transformation function. It is defined as inner product response in the high dimensional space as $K(x_u, x_v) = \phi(x_u)\phi(x_v)$.

$$\sum_{u=1}^{N} \alpha_u y_u = 0 \; and \; 0 \leq \alpha_u \leq C \tag{2}$$

The penalty parameter (C) controls the trade-off between errors of the SVM on training data and margin maximization. Each nonzero α_u value corresponds to a support vector. In the prediction phase, with using all support vectors (N_{sv}), nonlinear classification can be computed as in (3) for an unlabeled sample of x.

$$f = sgn\left(\sum_{u=1}^{N_{sv}} \alpha_u y_u K(x_u, x) + b \right) \tag{3}$$

In this study, radial basis function kernel (RBF) $K(x_u, x_v) = \exp(-\gamma \|x_u - x_v\|^2)$ is selected as kernel function. With using one-against-all (OAA) or one-against-one (OAO) strategies, multiclass problems can be realized by SVM's binary classification structure. In the study, the best parameter optimization of SVM classifier is realized with grid search method. Penalty parameter (C) is evaluated between [1-100] with a step size increment of 2. In RBF kernel, the best γ parameter is searched between [0.01-10] with a step size increment of 0.1.

4 Findings

First, the following features are selected and applied to the churned users data sets.

- x1: Days between first and last usage date
- x2: Actual usage time in minutes between first-last usage dates

- x3: Number of days app is used between first-last usage dates
- x4: Number of sessions opened between first-last usage dates

One of the interesting and unexpected results are seen when histogram for the first feature (x1) is examined. As seen in Figure 2, mode value of x1 is 0 which indicates that a significant number (20%) of the churned users decide to quit the application in their first day.

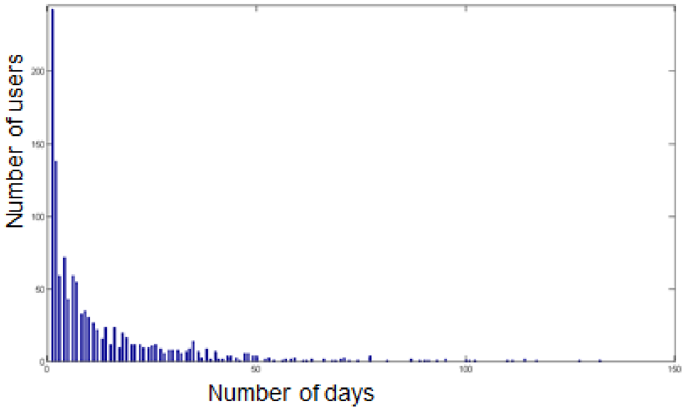

Fig. 2. Days between first and last usage date for churned users

After observing this fact, the churned user subset is divided into two following further subsets with the aim of finding a difference in usage patterns according to the first day usage.

- E1: churned users who have used the application only for their first usage day and then quit
- E7: churned users who have used the application more than seven days and then quit

Features x2 and x4 are being calculated for E1 and E7 subsets for the first day usage and the results given in Table 1 are obtained:

Table 1. First Day Usage

Defined Features	E1	E7
Average x_2 (minutes)	8.98	17.5
x_2 standard deviation (minutes)	12.66	26.77
Average of x_4 (number)	1.93	2.56
x_4 standard deviation (number)	2.02	3.42
x_2 median value (minutes)	7	8

According to the results, it is found that, 12% of churned users quit the application for using only one minute; on the other hand 11% of them quit the application after using 13 minutes. 225 of 242 E1 users (93%) used the application less than 40 minutes.

E1 and E7 users are being classified using Support Vector Machines (SVM) according to x2 and x4 features. E1 class is labeled with 0 and E7 class is labeled with 1, "cost" parameter and "gamma" parameter is selected to produce the best result and SVM model is constructed. Then, SVM which is considered to be a successful classifier produced 93% of separability. This result indicated that, users who may quit the application in their first day can be detected.

After finding that E1 and E7 users have different patterns, next question is to be able to find a churn pattern after the first day. Stickiness is considered to be an appropriate indicator of this pattern. It is expected that, churned users' stickiness converge to a level which is expected to be different (e.g. lower) than retained users. Stickiness is being measured as follows: x3/x1. In other words, stickiness equals to "the number of days app is used between first-last usage dates" divided by "days between first and last usage date". For example, a user uses the application for the first day has the maximum stickiness level of 1. If she/he does not use the application next day, stickiness level decreases to (1/2) and if she/he uses the application on the third day, stickiness level increases to (2/3). Users who continue using the application are labeled as N, and churned users (without a distinction of which day they used) are labeled as E.

Second usage date of users is extracted for E and N classes and it is found that E percentage of users who use the application:

- 1 day after first usage: 67%
- 2 days after first usage: 0%
- 3 days after first usage: 7%
- 4 days after first usage: 4%
- 5 days after first usage: 2%
- 6 days after first usage: 1%
- 7 days after first usage: 2%

For the N users the results are as follows:

- 1 day after first usage: 94%
- 2 days after first usage: 0.8%
- 3 days after first usage: 0.8%
- 4 days after first usage: 0.5%
- 5 days after first usage: 0.2%
- 6 days after first usage: 0.4%
- 7 days after first usage: 0.3%

In order to find, if there is a different pattern in terms of stickiness between E and N users; the SVM is applied to the following data set:

First Subset

Retained Users: used the application for January and February 2013

Churned Users: used the application in January but did not use in February 2013

Second Subset
Retained Users: used the application for February and March 2013
Churned Users: used the application in February but did not use in March 2013

Third Subset
Retained Users: used the application for March and April 2013
Churned Users: used the application in March but did not use in April 2013

Fourth Subset
Retained Users: used the application for April and May 2013
Churned Users: used the application in April but did not use in May 2013

The following six features are selected to apply machine learning algorithms to those four subsets between the first usage and last usage date of the pilot mobile application:

1. Total usage time (minutes)
2. Maximum usage time in a session (minutes)
3. Maximum usage time in a day (minutes)
4. Number of total sessions
5. Number of days application is used
6. Average daily usage time (minutes)

For example; On 25 March a user has opened three sessions with 9 minutes, 2 minutes and 4 minutes, on 26 March did not use the application and last usage happened on 27 March with a session time of 3 minutes. Then her/his features are calculated to be:

- Total usage time: (9+2+4+3=) 18 minutes
- Maximum usage time in a session: 9 minutes
- Maximum usage time in a day: 15 minutes
- Number of total sessions: 4
- Number of days application is used: 2 days
- Average daily usage time: (15+0+3)/3 = 5 minutes

The SVM results are found as follows:

- First subset is separable by 70%
- Second subset is separable by 75%
- First subset is separable by 71%
- First subset is separable by 69%

The results did not indicate a significant percentage of separation. It may be due to the time frame chosen. For example, users' usage time may span more than two months. Thus, a new and longer time frame is chosen for the next analysis. New time frame is chosen to be 7 months during November 2012 to May 2013. This data is again divided into two subsets for churned and retained users in order to find the differences in their patterns. After applying SVM, the results are obtained as follows:

- Number of churned users are 627 and retained users are 112. When the six features are taken and SVM is applied, it is found that these two data are separable by **91%** which is an acceptable level.
- In order to deepen the analysis, users who use the mobile application more than one month among the retained users are extracted. They are 48 people. And then the same SVM is used to the data set of these 48 people and 627 people. This time a better separation level is obtained: these two data are found to be separable by **98%**.
- Then, a new subset is taken for churned users. The churned users who have quit the application for their first day is excluded. New data set consist of 498 churned users who have used the application more than one day and 48 retained users who used the application more than one month. SVM is applied and these two data are found to be separable by **98%**. The same result is obtained with the previous result.

Since SVM produced successful results with this new data set (November 2012-May 2013), stickiness study is repeated with the same data set and the following findings are obtained:

- Stickiness values for the last usage day of 627 churned users and 112 retained users are calculated to be 0,65 and 0,46 respectively. SVM result indicated **84%** separability.
- Stickiness values for the last usage day of 48 retained users who used the application more than one month is calculated to be 0,14. SVM is applied to the data set of 48 retained users and 627 churned users where result indicated **92%** separability.
- Stickiness values for the last usage day of 498 churned users who used the application at least two days are found to be 0,56. SVM is applied to the data set of 48 retained users and 498 churned users where result indicated **91%** separability.

After obtaining successful results in forecasting a user's decision to churn or continue using the application, the next action would be, estimating the next usage date and time for the user. In order to estimate this information, "probability of usage" factor should be calculated and in order to do this calculation, one needs to know the causal relations that effect probability of usage. For this reason we constructed this causal relationship as depicted in Figure 3.

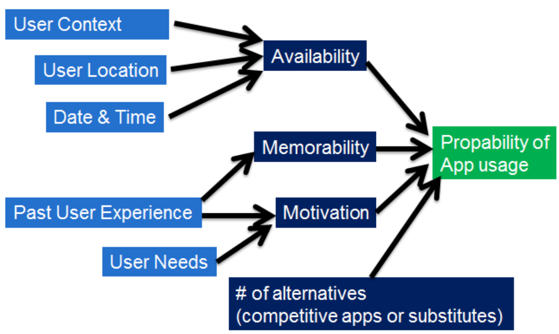

Fig. 3. Causal Factors Effecting Probability of App Usage

Availability is a necessary and precondition of usage but it is a not sufficient condition per se. Users use applications in their spare times. Availability detection is determined by several conditions such as user context (for example a meeting context is different than working alone at work environment, or if the user user may be ill, lying on bed in a work day where s/he can use apps to kill time), user location (work/home/outside/transportation; indoor/outdoor) and date & time (work day or holiday).

Memorability is another required condition of the user. (i.e. user considers using the app if and only if s/he remembers the existence of the app and includes using app in her/his alternatives in his/her spare time).

After user considers using the app, his/her decision is based upon her/his Motivation to use. (i.e. does she/he wants or needs to use the application?). User current need to use the application and her/his past experience with the application affects her/his motivation.

There is a negative correlation between the Number of Alternatives and application usage. (For example if there are lots of alternatives for the user then her/his probability of using the app decreases. This is why some people use those apps during public transportation).

Since the time needed to collect and classify data related with these causal factors, instead a statistical study is done in order to classify the usage time and their related context. After this study the following usage heat map of users are extracted as shown in Figure 4. The vertical axis indicates hour and horizontal axis indicates minutes.

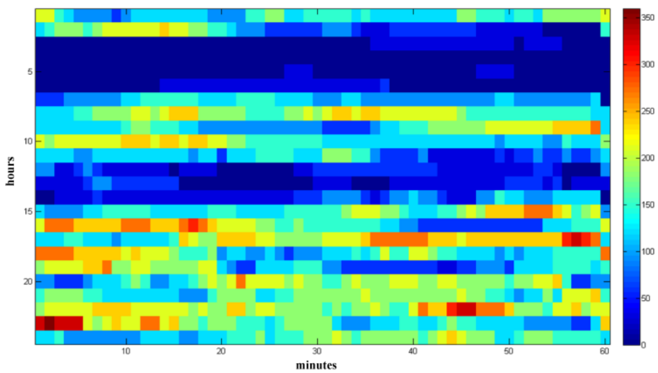

Fig. 4. Heat Map for App Usage Times

As seen in Fig. 4 heat map, usage intensity happens in three time frames. A sample of 30 users' location information is traced manually and possible context information is given for these time frames as listed in Table 2.

Table 2. Observed Context and Usage Intensity

TIME FRAME	HOURS	OBSERVED CONTEXT	USAGE INTENSITY
A	09:30 – 11:00	Outside	%13.38
B	16:00 – 19:00	Home, Work, Outside, Road	%19.33
C	21:00 – 02:00	Home	%37.36

In order to see the results of engagement and stickiness, push notification times are changed to these time frames and after sending one push notification for each time frame, the following results are obtained:

- Active user percentage increased from 1,72% to 2,08 % with an increase of nearly 21%.
- Total engagement (number of minutes application is used) increased 36%
- Engagement per user did not increase (slightly decreased 5%)

5 Conclusions and Future Work

This study was a continuation of a previous study where data were collected [4] and in this study they are used and findings are obtained. The following results are obtained from the study:

- In the first day usage, churned users are distinguished with 93% with the retained users.
- On the subsequent days, again churned users are distinguished with a 98% with the retained users.
- Three usage time frames that extracted are extracted are: 09:30-11:00; 16:00-19:00 and 21:00-02:00 and the context are observed to be outside; home/work/outside/road; home respectively. These three time frames consist of 70% of all usages.
- When the push notification time is change to those time frames, active user percentage and total engagement increased 21% and 36% respectively. However, engagement per users decreased slightly (5%).

As a future work, this study can be extended or related with the following objectives:

- The motivation of a user while using the application can be known
- The mastery level of the user can be estimated
- The payment trend of the user can be defined, so the user can be encouraged with rewards and new items

References

1. Apple Press Info, http://www.apple.com/pr/library/2013/10/23/Apple-Announces-iPad-Air-Dramatically-Thinner-Lighter-More-Power-ful-iPad.html
2. Android's Google Play Beats App Store, http://www.phonearena.com/news/Androids-Google-Play-beats-App-Store-with-over-1-million-on-apps-now-officially-largest_id45680
3. Microsoft by the Numbers, http://www.microsoft.com/en-us/news/bythenumbers/index.html
4. Gençer, M., Bilgin, G., Zan, Ö., Voyvodaoğlu, T.: A New Framework for Increasing User Engagement in Mobile Applications Using Machine Learning Techniques. In: Marcus, A. (ed.) DUXU/HCII 2013, Part IV. LNCS, vol. 8015, pp. 651–659. Springer, Heidelberg (2013)
5. Van den Poel, D., Larivière, B.: Customer Attrition Analysis for Financial Services Using Proportional Hazard Models. European Journal of Operational Research 157(1), 196–217 (2004)
6. Au, T., Guangqin, M., Shaomin, L.: Applying And Evaluating Models To Predict Customer Attrition Using Data Mining Techniques. Journal of Comparative International Management 6(1) (2003)
7. Burez, J., Van den Poel, D.: CRM At A Pay-Tv Company: Using Analytical Models To Reduce Customer Attrition By Targeted Marketing For Subscription Services. Expert Systems with Applications 32(2), 277–288 (2007)
8. Hung, S.-Y., Yen, D.C., Wang, H.-Y.: Applying data mining to telecom churn management. Expert Systems with Applications 31(3), 515–524 (2006)
9. Coussement, K., Van den Poel, D.: Improving Customer Attrition Prediction By Integrating Emotions From Client/Company Interaction Emails And Evaluating Multiple Classifiers. Expert Systems with Applications, vol. 36(3(Pt. 2)), Pt. 2, pp. 6127-6134 (2009)
10. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2007)
11. Hu, X.: Data Mining Approach For Retailing Bank Customer Attrition Analysis. Applied Intelligence 22(1), 47–60 (2005)
12. Wu, R.-C., Chen, R.-S., Chen, C.-C.C.J.Y.: Data Mining Application In Customer Relationship Management of Credit Card Business. In: IEEE Computer Software and Applications Conference, vol. 2 (2005)
13. Hoggart, C.J., Griffin, J.E.: A Bayesian Partition Model For Customer Attrition. In: Bayesian Methods with Applications to Science, Policy and Official Statistics (Selected Papers from The Sixth World Meeting of the International Society for Bayesian Analysis), pp. 223–232 (2001)
14. Schölkopf, B., Smola, A.: Learning with Kernels. MIT Press (2002)
15. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons (2001)