Epileptic Seizure Detection Using a Wrist-Worn Triaxial Accelerometer

Jussa Klapuri

Preface and Acknowledgements

I wish to thank first and foremost my supervisor Professor Samuli Siltanen and Anna Kaisa Kylmä from Vivago for their contribution as being my instructors during this thesis process. I would also like to thank Tuomas Nikkonen from the Inverse Problems Group and Eelis Cederström from Vivago for their ideas and our discussions about the topics related to this thesis. Additionally, thanks to Eeva-Liisa Metsähonkala and other people at HUS for providing me with the epilepsy dataset, and to Vivago for lending me their product to help me record my own datasets. I am particularly grateful to my two test subjects for being willing to record data and keep a diary.

Special thanks to Kacsóh family, especially Jaana and László, for indirectly influencing me to pursue greater challenges in my academic career. Without their contribution, the thought of studying towards a second degree might not have crossed my mind in early 2010.

For their excellent contribution in proofreading, I wish to thank Taneli Riitaoja and Mika Göös. Finally, thanks to my family and close ones, especially Ella, for supporting me through this thesis project (again).

Espoo, 1st November, 2013
Jussa Klapuri

Contents

List of Abbreviations

ACM	Accelerometry
ANN	Artificial Neural Network
AR	Autoregressive
DCT	Discrete Cosine Transform
DPS	DCT-PCA-SVM-Model
EEG	Electroencephalography: the recording of electrical activity along the scalp
FFT	Fast Fourier Transform
GTC	Generalized Tonic–Clonic (Seizures)
HMM	Hidden Markov Model
PCA	Principal Component Analysis
SVM	Support Vector Machine

Chapter 1

Introduction

Epilepsy is the cause of a significant portion of the world's disease burden. Its economic, social and personal costs are due largely to uncontrolled seizures (Pitkänen et al., 2005). It is estimated that almost one percent of the world population suffers from the effects of epilepsy. In the United States alone, the annual costs for the estimated 2.3 million prevalent cases is estimated at $12.5 billion. The direct costs of epilepsy account for 15%, including diagnosing and medical resources, and the rest 85% account for the indirect causes, such as foregone earnings and reductions in household activities because of epilepsy-related morbidity and mortality. (Begley et al., 2000) In light of these facts, any effort to mitigate the effects of epilepsy is well-motivated.

Seizure detection generally describes the instantaneous detection of an epileptic seizure that triggers an alarm system intended to alert for right assistance in situations that require immediate intervention (Nijsen et al., 2005). In addition, seizure detection also provides long-term benefits, because the information collected on the different seizure types, frequencies and their distribution can lead to better management of daily care and better titration of antiepileptic drugs. In the scope of this thesis, seizure detection generally means simply detecting a seizure, not reacting to it.

The "gold standard" for seizure detection is the recording of electrical activity along the scalp, i.e., electroencephalography (EEG), combined with video monitoring in a laboratory environment (Nijsen et al., 2005). While an approach such as this obviously results in large amounts of useful and accurate data, the data is very costly and this kind of measurement system cannot be considered as a realistic environment for everyday life. Physicians make treatment decisions on epilepsy mainly based on seizure frequency observed during monitoring phases generally lasting less than a week. Hence, the effectiveness of a given pharmacological treatment is typically based on its impact on the frequency of seizure events over an extended period of time. (Deckers et al., 2003) In practice, the impact is assessed by the patient or caregivers quantifying seizure frequency, yet both of these

measures can be inaccurate and misleading. Thus, there is a need for a non-invasive device capable of measuring seizure frequency over extended periods of time with high *sensitivity* (rate of correctly identified seizures) and *specificity* (rate of correctly identified nonseizures). High specificity is also often referred to as having a low amount of false alarms.

In this thesis, the research question is:

> *How well is it possible to detect severe epilepsy seizures solely by a triaxial accelerometer embedded into a watch, and what are the mathematical or statistical models needed to facilitate such seizure detection.*

Accelerometry (ACM) helps in solving this problem by providing rather simple and cheap sensors that can easily be worn even 24 hours a day. In this thesis, ACM will play a central role in contrast to video-EEG monitoring. A triaxial accelerometer is a device that measures acceleration forces, i.e., g-forces, in all three dimensions. This means that movements and their directions will be detected according to the frame of reference. Without a gyroscope, however, the orientation of the sensor is not explicitly known, although for a stationary sensor it can easily be approximated by computing the gravity acceleration vector (Degen et al., 2003).

Accelerometers generally differ by their sampling frequency, resolution and maximum g-forces measured. In this thesis, a Vivago Wellbeing Watch was used for recording ACM data. The watch is capable of recording 1 byte of acceleration data on 3 channels using 40 Hz sampling frequency and measuring up to ± 8 g of acceleration. The maximum g–force a human body can produce in a very heavy physical activity is around 12 g (Bouten et al., 1997), but in everyday life, and in seizure detection, measuring at most 8 g is enough by far. For comparison, a Nintendo Wii Remote measures ± 3 g at 100 Hz and has been used in an epilepsy seizure study as the principal accelerometer fixed onto a patient's forearm with promising results (Schulc et al., 2011). There has also been at least one study where recognition accuracy as a function of the accelerometer sampling frequency was studied (Maurer et al., 2006). The conclusion was that no significant gain in accuracy is achieved by using a sampling frequency above 20 Hz or being able to measure amplitudes over ± 2 g.

A model for the seizure detection must be able to learn from the accelerometry data, involving seizures, and to generalize to new and previously unseen ACM data. The model must be reasonably quick in recognizing seizures that are currently happening, or starting to happen. In this particular case, the model should be able to disregard the orientation of the watch and the choice of the wrist.

We approach this challenge by utilizing four kinds of accelerometry datasets recorded by a single wrist-worn accelerometer by three different test subjects. The research is conducted by taking the following steps:

1. Producing an **artificial seizure training set** by a test subject mimicking symptoms of severe epilepsy seizures.

2. Producing an **artificial daily life training set** by recording a test subject's movements during approximately one day's worth of data including many different kinds of nonseizure movements (e.g. sprinting and cooking) and movements that slightly resemble seizure-like movements (brushing teeth, washing hands).

3. Producing an **artificial daily life validation set** by a second test subject that includes some simulated seizures.

4. Producing an **artificial daily life test set** by a third test subject that includes some simulated seizures.

5. Training a model for seizure detection by using the **training sets** and fine-tuning the model parameters such that the seizure detection accuracy on **validation set** is maximized.

6. Testing the seizure detection accuracy of the model on the **test set** to measure the generalization error of the model.

7. Testing the model on the poorly labeled **real world epilepsy dataset** to study if any known epilepsy seizures are detected.

In this thesis, the epilepsy seizure detection is implemented by splitting the ACM signal into 5 s windows and then using discrete cosine transform (DCT) and principal component analysis (PCA)for preprocessing and feature extraction. Finally, the 5 s windows are classified using support vector machines (SVM). These are the principal steps of the *DPS* model, which is introduced in Chapter 4.

The performance of the model on the validation set will help to measure which parameters and which kind of training data work the best for this specific case of seizure detection. Additionally, good performance on the real world epilepsy dataset will give more certainty on how the model would perform in real life. However, it is not known whether the known seizures in the epilepsy dataset actually contain wrist movement and thus the results on this dataset are not fully representative of the potential of the DPS model.

The structure of this thesis is as follows: Chapter 2 includes the literature review related to activity recognition, seizure detection, epileptic seizures and some approaches often used in modeling. Chapter 3 contains in-depth analysis of the utilized datasets and Chapter 4 mathematically defines the preprocessing, feature extraction and classification phases for the seizure detection functionality of the DPS model. Chapter 5 describes how

the training and test phases were implemented and Chapter 6 concludes the results and discusses future work.

Chapter 2

Activity Recognition and Seizure Detection

2.1 Activity Recognition

Activity recognition is a technique which can recognize human activities or gestures via the help of computer systems using signals and data obtained from sources such as electromyography (electrical activity of skeletal muscles), audio & image sensors, and accelerometers. The results are applied in many areas including biomedical engineering, medical nursing and interactive entertainment. (Yang et al., 2008) The research questions in activity recognition studies are mostly concerned with trying to recognize between 3 to 20 different activities as well as possible. Some of the studies are set in an artificially constrained laboratory setting (Yang et al., 2008), while others use data recorded in the real world and labeled by the person providing the data (Bao and Intille, 2004; Stikic and Schiele, 2009). Often the number of people or patients involved in these studies range from 2 to 20 and the data is recorded during the span of several days. From purely machine learning perspective, having more test subjects perform the same moves helps in *generalizing* the recognition model in a way that similar results can be expected from new test subjects. Generally the activity recognition accuracy has been around 84–97% with some specific activities being recognized with even higher accuracy, such as "Kung Fu arm movements" and dish washing (Bao and Intille, 2004).

Another field closely related is gesture recognition, which can often be implemented through video feed, such as Xbox Kinect (Microsoft, 2013). There have also been studies in regards to accelerometers used in hand gesture recognition with gestures similar to ones used in video games (Moiz et al., 2011). While these kinds of studies may not seem to relate to epilepsy detection, many of the same features computed from the ACM data and the choice of classification models are applicable in both fields.

2.2 Seizure Detection

This section describes the detection of epileptic seizures and reviews many studies related to it. Seizures can be divided into different types, and it is important to first comprehend this taxonomy to better understand what seizures different detection algorithms are attempting to detect.

2.2.1 Epileptic Seizures

There are about 40 different types of epileptic seizures, which can generally be divided into *partial seizures* and *generalized seizures*. Partial seizures begin in a part of one hemisphere of the brain while generalized seizures begin in both hemispheres at the same time. Partial seizures can be further devided into *complex* and *simple* seizures, depending on whether consciousness is impaired or not. Partial seizures may involve motor signs, but generally do not. These kinds of seizures without motor signs are often called *auras* and cannot be detected with ACM. However, partial seizures can evolve into generalized seizures which include the most notable seizure types that are the foci of this thesis. Typical classification of generalized seizures is:

1. Absence seizures
2. Myoclonic seizures
3. Clonic seizures
4. Tonic seizures
5. Tonic-clonic seizures
6. Atonic seizures

The *convulsive type* seizures include tonic-clonic seizures (*grand mal* seizures), purely clonic seizures, purely tonic seizures and various combinations of these manifestations. The nonconvulsive seizures include typical absence seizures (*petit mal* seizures) and atypical absences. Typical absences involve brief losses of consciousness while atypical absences last longer. Atypical seizures also include some sort of physical movements, such as a brief head nod or jerking of the shoulders.

Tonic-clonic seizures are seizures that most people think of as epilepsy. At the start of the seizure the person goes through a tonic phase where the person becomes unconscious and the body becomes rigid and may fall if standing. After which follows the clonic phase, consisting of spasms and rhytmical convulsions. During this time the skin color may become very pale or bluish. After the seizure, when the convulsions stop, the person may feel tired, confused, have a headache or have a need to sleep.

In a tonic seizure the person's muscles suddenly become stiff, which often causes the person to fall down if standing (usually backwards), with the risk of injuring one's head.

Atonic seizures involve sudden relaxation of muscles, causing a person to fall (usually forwards), which may injure one's head and face. Both tonic and atonic seizures happen without warning and tend to be very brief.

Myoclonic seizure is a kind of "muscle jerk" that usually affects arms or legs, but can also affect top half of the body or head. Finally, the clonic seizures are convulsive seizures where the person's body does not go stiff at the start as in tonic-clonic-seizure.

Myoclonic, atonic, and brief clonic seizures can all result in falls, referred to as *astatic seizures* or drop attacks. While myoclonic seizures are brief, they often happen in clusters and the person is conscious. However, the person suffering from myoclonic seizures may often have other seizures, such as tonic-clonic seizures. (Epilepsy Society, 2013; Pitkänen et al., 2005)

In short, the biggest difficulties in recognizing epilepsy seizures relate to the multitude of possible seizure types as well as seizures being quite individual. Also, most seizures happen suddenly without warning, last a short time (from seconds to few minutes) and stop by themselves. In theory, the types of seizures that may be possible to detect by using an ACM sensor are all generalized seizures excluding absence seizures, if the data is very well labeled. Since this is rarely the case, only the most physical types of seizures can generally be recognized from the data, meaning clonic and tonic-clonic seizures. Astatic seizures, or falls, give a specific ACM signal but this is quite difficult to reproduce accurately and safely at the same time. Within the epilepsy dataset used in this thesis, no known astatic seizures happened.

2.2.2 Epileptic Seizure Detection Studies

A decision one has to make first in a seizure detection study is the number and placement of sensors to be worn. The outcome of this decision may cause usability issues outside the laboratory environment. To counter this problem, the easiest or the most natural place for a sensor to be worn has to be the wrist or a pocket, in other words, a watch or a mobile phone. In some studies, the wrist has been concluded to be an unapplicable site for ACM, e.g. for fall detection (Kangas et al., 2008; Doughty et al., 2000). In many studies, the number of accelerometers is between 1 and 5, often one accelerometer in the chest and one in each limb (Bao and Intille, 2004; Nijsen et al., 2010). With 5 accelerometers, seizure detection should become somewhat easier since the seizure movements can be distinguished from simple hand and wrist movements. With only one accelerometer attached to a wrist, the problem is that the sensor can move by itself in ways that might resemble seizures, such as brushing teeth (Beniczky et al., 2013). There is also a study of sensor displacement, i.e., how does recognition suffer when the sensors get displaced because of physical activities or wearing them for extended periods of time (Förster et al., 2009). When a wrist watch is tightened well enough, an accelerometer inside the watch

is not affected by displacement.

The most interesting studies related to the topic of this thesis are those that attempt to detect seizures with exactly one accelerometer attached to the wrist. The master's thesis of Joachim Elevant (Elevant, 1999) is titled "Monitoring Epilepsy With a Wrist carried Motion Sensor" and at first appears to be very relevant, but the approach to the problem seems to be from an electrical engineer's perspective. It involves simple bandpass filters and is very patient-specific, requiring specific cutoff frequencies for each patient.

Most thorough studies in seizure detection have involved Video-EEG monitoring coupled with one or more accelerometers (Beniczky et al., 2013; Nijsen et al., 2010; Schulc et al., 2011). To enable Video-EEG monitoring, the subjects had to be located in an artificially constrained laboratory setting for some hours or days. There exist many EEG-based seizure detection algorithms implemented in many inpatient epilepsy monitoring units. However, only a few patients are willing to wear EEG electrodes for signal acquisition on long-term basis (Schulze-Bonhage et al., 2010). This is why it would be better if patients could be monitored with confidence by using ACM. To give an idea of the scale of some of the studies, Beniczky et al. (2013) gathered an unparalleled 4878 hours of data in total with a wristwatch accelerometer. Majority of the data also included video-EEG data since the reference standard for seizure times were identified by experienced clinical neurophysiologists. While the results were good and 89.7% of all seizures were detected, the actual accelerometer used was practically a black box, designed by Danish Care Technology ApS. This means that no algorithms were given or speculated in the study. Another study involving epileptic seizure detection using a wrist accelerometer as a black box is by Lockman et al. (2011). They managed to detect seven of the eight GTC seizures occuring during the monitoring phase of six different patients while raising 204 false alarms in the process, which is a very high number. Again, the way to verify the actual seizures was by measuring Video-EEG alongside ACM data. The accelerometer they used was called SmartWatch by Smart Monitor, Inc. While (Beniczky et al., 2013) and (Lockman et al., 2011) are very interesting studies for epileptic seizure detection, they offer absolutely nothing of value for the implementation of a seizure detection algorithm because of using black box models in detection.

The experiments by Schulc et al. (2011) included a single wrist-worn accelerometer supported by Video-EEG monitoring without using a black box device. For epileptic seizure detection, they used a rather simple heuristic model computing mean, standard deviation and intensity (energy) of a signal with proper cutoff values. In the end, they achieved specificity of over 88% when sensitivity is set to 100% which at first seems to be a great result. However, the data consisted of only four GTC seizures by three people out of twenty people monitored. The obvious conclusion is that they will need more data in order to significantly confirm the positive results. There has also been some research done in myoclonic seizure detection, which is especially interesting since severe tonic–

clonic seizures are often preceded by myoclonic seizures. This means that detecting these myoclonic seizures efficiently might give an early warning (Nijsen et al., 2010), which might be the relevant next step after being able to detect the actual GTC seizures.

It is worth noting here that fall detection is another field that is related to seizure detection. Fall detection has applications in healthcare for the elderly people (Doughty et al., 2000; Kangas et al., 2008), and in epilepsy detection, especially astatic seizures, but this presents a problem of gathering real world data and in order to protect the patients' safety, no known drop attacks happened in any of the studies referenced here. Then again, falling data resembling real world cases could be manufactured quite easily with proper safety precautions.

2.3 Models in Literature

For quite abstract ACM data where only a couple of classes are used in classification, the feature extraction phase gains more weight. The feature extraction from acceleration signals is generally done by taking a window of 128, 256 or more samples. These windows are often overlapping each other by 50% so that the next window starts in the middle of the current window. The time domain features that are generally computed for each window are: mean, standard deviation, energy, and correlation between signals (Ravi et al., 2005). Additional features include autoregressive (AR) coefficients (He and Jin, 2008), fast Fourier transform coefficients (Nijsen et al., 2006), tilt angle and signal-magnitude area (SMA) (Khan et al., 2010). Finally, discrete cosine transform coefficients were utilized by He and Jin (2008) with great success and this technique will be discussed in Chapter 4.

The models used for classifying activities and seizures are quite varied, but there are some general trends. In general, the results from different studies may not be fully comparable, because the test settings can vastly differ and the number of patients or seizures analyzed may be so small that the statistical uncertainty is rather high. The earlier work gave more focus on simpler methods like decision trees and naive Bayes, with reasonable overall accuracy rate of 84% (Bao and Intille, 2004). Later on more advanced methods have improved the accuracy rate along with less false positives. One way is to apply hidden Markov models (HMM) with Bayesian analysis, which has given promising results (Jallon, 2010). Artificial neural networks (ANN) have also been implemented using time delay features, thereby improving the results (Moiz et al., 2011). Conditional random fields (CRFs) are especially interesting due to their temporal classification properties and generally outperforming HMMs (Lee et al., 2011; Vail et al., 2007). In the early phases of this thesis, CRF approach was considered to be the most likely approach to be taken. Also, the AR features seem to work well with ANNs (Khan et al., 2010), and support

vector machines (SVM) (He and Jin, 2008). In fact, the latter study claims that AR features work better than the traditional widely used time domain features combined with FFT features. The same authors, He and Jin, later released a paper that used SVM for classification, but instead of AR features, utilized DCT features with principal component analysis (PCA) (He and Jin, 2008). This approach is described in detail in Chapter 4.

There are also many studies using some heuristic models in such a way that the process cannot be referred to as machine learning. For example, Cuppens et al. (2009) use data produced by 5 accelerometers and attempt to detect nocturnal frontal lobe seizures by removing noise and applying a simple moving average filter. This resulted in sensitivity of 91.67% and specificity of 83.92%. Bersch et al. (2011) only used visual inspection of the signals in both time and frequency domain to distinguish between falling, sitting, and walking. The conclusion was that it is not possible to distinguish sitting from falling by using a single accelerometer in a watch.

Chapter 3

Datasets

This thesis focuses on 4 kinds of ACM datasets: artificial training dataset, artificial daily life validation and test set, and real world epilepsy dataset. The reason for so many datasets is that in the original epilepsy dataset, which contained many days of acceleration data by several patients, the labeling of the seizures was poor. This means that without any additional video data of the subjects to help labeling, the exact time and length of the seizures are not known and thus supervised learning is impossible.

The principal reason for producing the artificial datasets was to test the model on strictly labeled datasets. All data was gathered by a Vivago watch that measures tri-axial acceleration with 40 Hz sampling frequency and is able to measure up to $\pm 8\,\mathrm{g}$ of acceleration. In practice, one sample of data contains three signed integers of one byte (value between -128 and +127) and there are 40 such samples in a single second. All of the artificial data was produced by 3 test subjects, 2 males and 1 female, doing various activities and occasionally performing the following gesture:

Trembling and rhytmical convulsions of the hands and possible the body (2–5 Hz).

At first the plan was to mimic different types of movements that seem to often happen during a GTC seizure, but since the irregularity of these movements was high, only the trembling kind was performed. In further studies with a video monitoring capability, including different kinds of movements would be a more realistic option.

3.1 Artificial Seizure Data

The objective of the artificial seizure dataset was to mimic hand gestures during a tonic-clonic seizure, i.e., produce data similar to what an accelerometer would measure during a real tonic-clonic attack. For this reason, several videos of tonic-clonic seizures were

analyzed (Panayiotopoulos, 2006). Important aspect to note here is that the length of the datasets is not as important as variety. Test subjects wore the watch in their dominant hand, contrary to the real world epilepsy dataset. This was done because most of the interesting activities are done using the dominant hand, such as brushing teeth or playing guitar and thus the false positives (nonseizures) would be easier to detect as such. In a real life situation, many of the false positives would be ruled out if the epilepsy patient wore the watch on the non-dominant hand. On the contrary, the epileptic seizures would still be detected with the same accuracy, assuming that the seizures affect both hands in the same way.

The artificial seizure training set was in essence very short, consisting of only two minutes worth of seizure-like movements while sitting and laying down. Special care was given to the position of the wrist such that several orientations of the watch were recorded while acting out a seizure. It is worth noting that Lockman et al. (2011) studied 40 patients who had a total of 8 GTC seizures. During the monitoring period, 204 additional nonseizures were detected, i.e., false positives. The nonseizures typically consisted of various normal, rhytmic movements such as:

- Wiping off shirt
- Shaking a bottle
- Fanning with a card
- Brushing teeth
- Exercising
- Arranging bed
- Nail filing
- Changing watch
- Washing up
- Tapping pen
- Moving arm up and down

These are the kinds of nonseizure movements that might resemble a seizure.

The artificial daily life training set was produced by the author by combining different kinds of ACM datasets gathered during the summer that included ordinary daily and sports activities such as:

- Walking
- Cooking
- Driving
- Writing
- Typing
- Running
- Sprinting

- Running up and down stairs
- Exercising on a mat
- Running on a cross trainer
- Playing disc golf
- Washing hands
- Playing computer games using a mouse

These were marked as not being seizures and in total the size of the data matrix was 2284467×3, which is roughly 16 hours of data.

3.2 Artificial Daily Life Validation and Test Set

The test subjects that produced the validation and test sets were informed to perform 10–20 seizures during a 1–2 day period in different positions, i.e., laying down or sitting on a chair. Also it was encouraged not to have too many seizures in too short order so that the difference between seizures might be affected by such real-life circumstances as time of the day, awareness, possible physical exhaustion and physical location. Also, having two seizures with too little time inbetween might indicate that there was only one seizure, of which the middle part was not recognized for whatever reason. Test subjects were instructed by showing a video of a real GTC seizure (Struska, 2013) and instructed to mimic the trembling or the shaking part of it. This meant that since their movement was not a real seizure, but their interpretation of it, detection of this movement by the model would reinforce the validity of the generalization property.

Both subjects were either sitting or laying down and the exact position of the hand varied from touching the body to being clearly apart. The exact position and the type of movements differed somewhat between the subjects, which meant that the data produced was more generalized than just acceleration data from a very specific movement, such as hitting a golf ball.

The validation set consisted of approximately 42 hours and 29 minutes of data, or 6192299×3 samples. During this time the subject had 7 seizures and wrote down some other activities such as brushing teeth, cleaning up and taking care of dogs in a kennel.

Test set was recorded during a 24 hour and 15 minute period with 3534336×3 samples and consisted of 13 seizures. Other activities written down by the test subject included: washing dishes, brushing teeth, jogging and playing guitar.

3.3 Epilepsy Data

Real world epilepsy dataset was provided by Vivago in collaboration with Eeva-Liisa Metsähonkala from HUS, The Hospital District of Helsinki and Uusimaa (Metsähonkala,

2012). The dataset contained data from 6 different patients having a total of 20 seizures of which most should be motor seizures. The datasets were recorded by having the patients wear the accelerometer watch for 1 to 4 days in a laboratory setting. There were nurses observing most of the time and they marked down the starting time of the seizures and other anomalies, such as stopping the recording if there was a need. Out of 6 patients, 5 wore the watch on the left wrist and one on the right wrist. The properties of the epilepsy datasets are shown in Table 3.1.

Table 3.1: Properties of the epilepsy datasets.

Dataset	Length (labeled)	Length difference in ACM data	Seizures labeled	Wrist
L1002_04	4 d 0 h 8 m	-35 m	7	Left
L1002_13	3 d 16 h 57 m	-35 m	4	Left
L1002_28	3 d 18 h 31 m	-29 m	1	Left
L1004_16	2 d 19 h 20 m	-1 h 11 m	1	Left
L1008_21	4 d 0 h 12 m	+2 h 54 m	4	Right
L1011_16	2 d 2 h 14 m	-6 h 45 m	3	Left

The main challenge with analysing this dataset was that the number of data points collected at 40 Hz does not match the times noted down by nurses. For example, if the notes say the recording had been going for 4 days and 8 minutes, the data produced by the watch may indicate it had recorded 35 minutes less. The differences between the time observed in real life and the time recorded in the data differed between each patient so it was not simply a matter of accelerometers sampling frequency being slightly different than exactly 40 Hz. Thus it was not determined whether the recording had been started later than indicated, ended earlier, sampling frequency fluctuating or differing from 40 Hz, or all of these together. Also the actual labeling caused some mistrust in the actual timestamps of the seizures, since for one test subject the first seizure was labeled as starting before the watch was turned on. It would have helped if there was Video-EEG data also available. For these reasons, labeling the ACM data beforehand was practically impossible because there was not even knowledge on how would a seizure look in the data, only some general properties like the periodicity of trembling.

Chapter 4

Models

This chapter defines the three major techniques used in epilepsy seizure detection model of this thesis, namely the discrete cosine transform (DCT), principal component analysis (PCA) and support vector machine (SVM). This model is referred to as DPS model and all the constituent techniques are applied as described in He and Jin (2009). The process of epilepsy recognition is visualized in Figure 4.1, where DCT and PCA are considered preprocessing methods and SVM the actual classifier between epilepsy events and normal events based on an accelerometer signal.

The notation in this chapter attempts to be consistent between the models. In general, scalars are denoted by a single symbol like x, while vectors are denoted by a boldface symbol $\boldsymbol{x}$. Matrices are denoted by bold uppercase symbols, e.g., $\boldsymbol{M}$. In general $\boldsymbol{x}$ represents a training set sample and $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_n]^T$ represents the Lagrange multiplier vector.

4.1 Discrete Cosine Transform

Ahmed et al. (1974) introduced the discrete cosine transform in the early seventies and it has since been applied in various fields, mostly in signal and image processing. The popularity of DCT is due to its "energy compaction" property: DCT concentrates most of the energy in a signal in a few coefficients enabling to compress a signal and losing only very little of the information contained in the signal. Similar kind of compression technique, using discrete wavelet transform, has been successfully applied to the acceleration data produced by Vivago Wellbeing Watch in Pekko Tuominen's bachelor's thesis (Tuominen, 2013). In everyday life, one of the most well-known applications of DCT is the lossy JPEG image compression algorithm, where 2-dimensional DCT is utilized as the integral part of the algorithm (O'Brien, 2005).

Let $x(n), n = 0, 1, \ldots, (N-1)$ represent a sequence of N samples. The discrete cosine

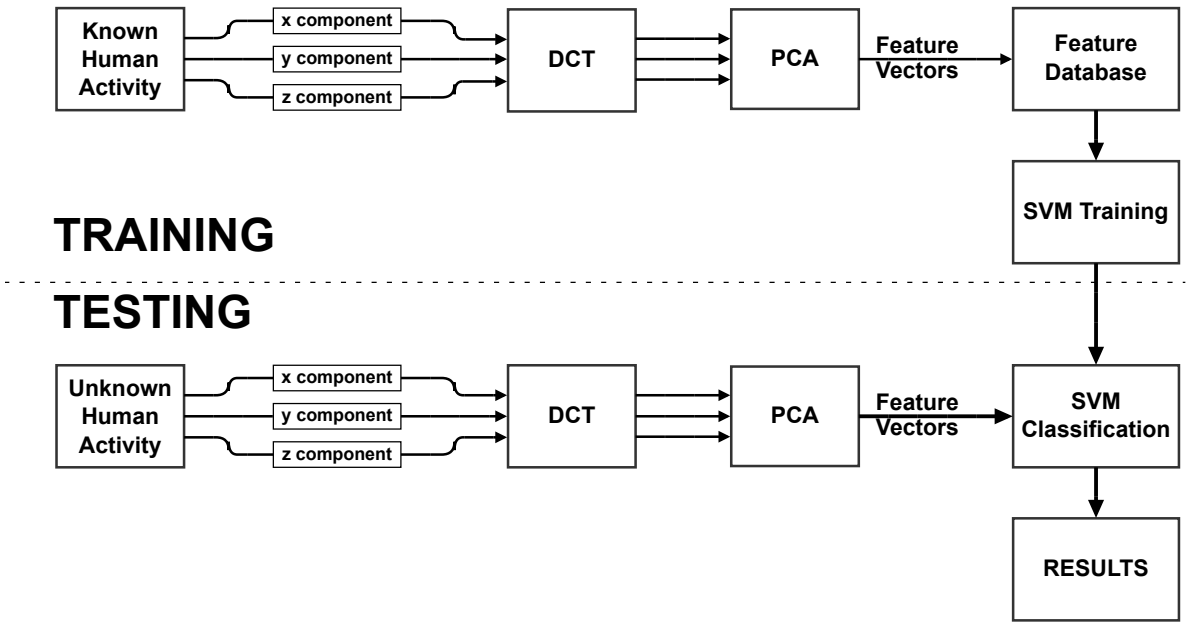


Figure 4.1: Diagram of the different phases in training and testing the DPS model used for epileptic seizure detection.

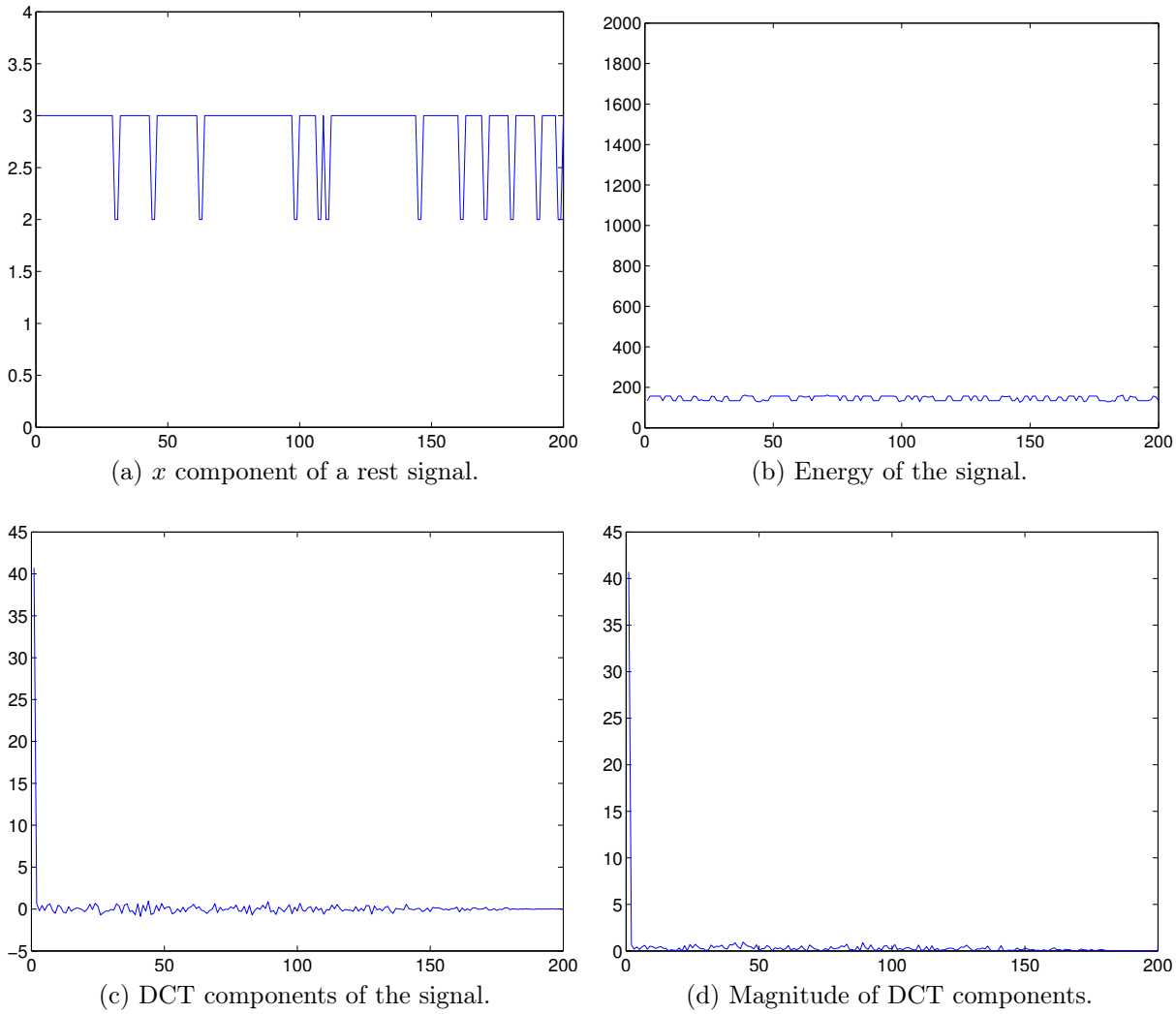
transform is then defined as

$$y(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad k = 0, 1, \ldots, N-1 \tag{4.1}$$

where

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}}, & k = 0 \\ \sqrt{\frac{2}{N}}, & k \neq 0 \end{cases}. \tag{4.2}$$

The time complexity of DCT algorithm can be reduced to $O(N \log_2 N)$ via a fast method similar to fast Fourier transform (Theodoridis and Koutroumbas, 2006). In Figure 4.2 it is shown that the DCT spectrum is very flat when applied to a rest signal, except the first offset value that is apparently representing the gravity vector (see Section 5.1). When the signal represents a seizure, the transformed signal looks very different, as in Figure 4.3. There is clearly a large spike near the 50th component that seemingly contains lot of information of the signal, compared to the rest of the spectrum.



(a) x component of a rest signal.

(b) Energy of the signal.

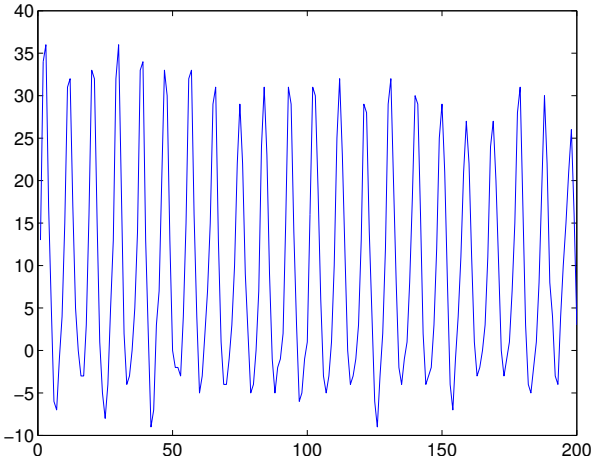
(c) DCT components of the signal.

(d) Magnitude of DCT components.

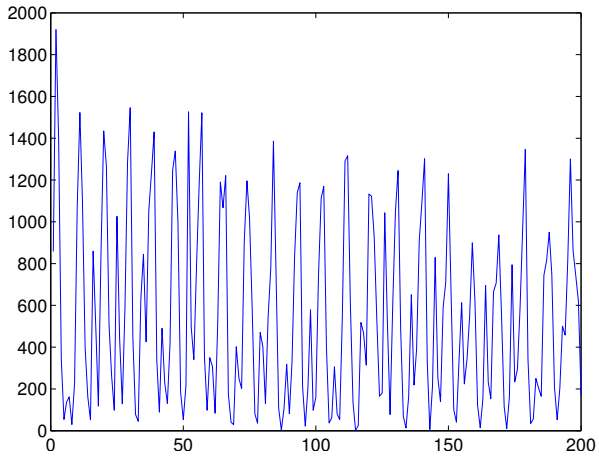
Figure 4.2: Visualizations of an accelerometer signal while laying completely still. Notice very homogenous DCT components in Figure (d).

4.2 Principal Component Analysis

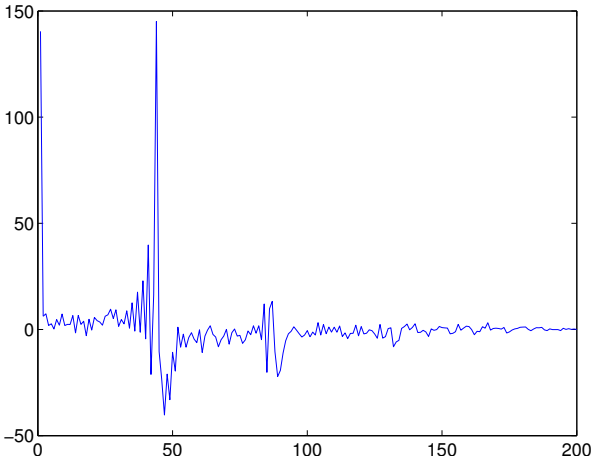
One of the fundamental steps in preprocessing is to lower the dimensionality of a vector by extracting the most important data features. This may improve the recognition process by considering only the most important, possibly uncorrelated, elements that retain maximum information about the original data with possibly better generalization abilities (Cios et al., 1998).



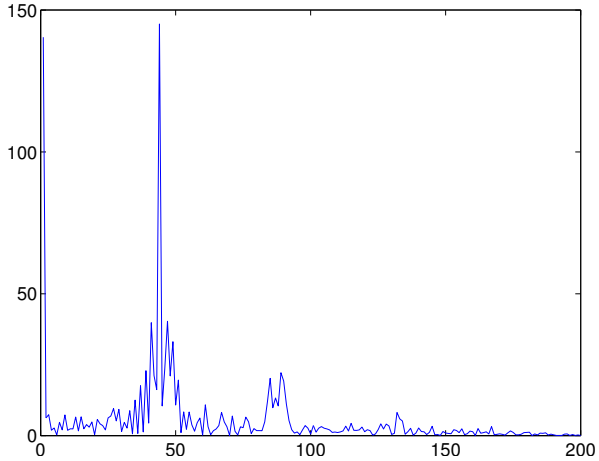
(a) x component of a signal representing trembling.



(b) Energy of the signal.



(c) DCT components of the signal.



(d) Magnitude of DCT components.

Figure 4.3: Visualizations of an accelerometer signal while trembling heavily. Notice very striking features in DCT components of Figures (c) and (d).

Principal Component Analysis, related to Karhunen–Loève transform, is a technique that can be used to compress high dimensional vectors into lower dimensional ones and has been extensively covered in literature (See e.g. Jolliffe (2002); Theodoridis and Koutroumbas (2006)). With two-dimensional data, PCA finds the orthogonal vectors representing the directions of most variance as illustrated in Figure 4.4. When using data produced by triaxial accelerometers, PCA offers a very useful method of compressing data and finding the most important components from input vectors that are constructed by adding

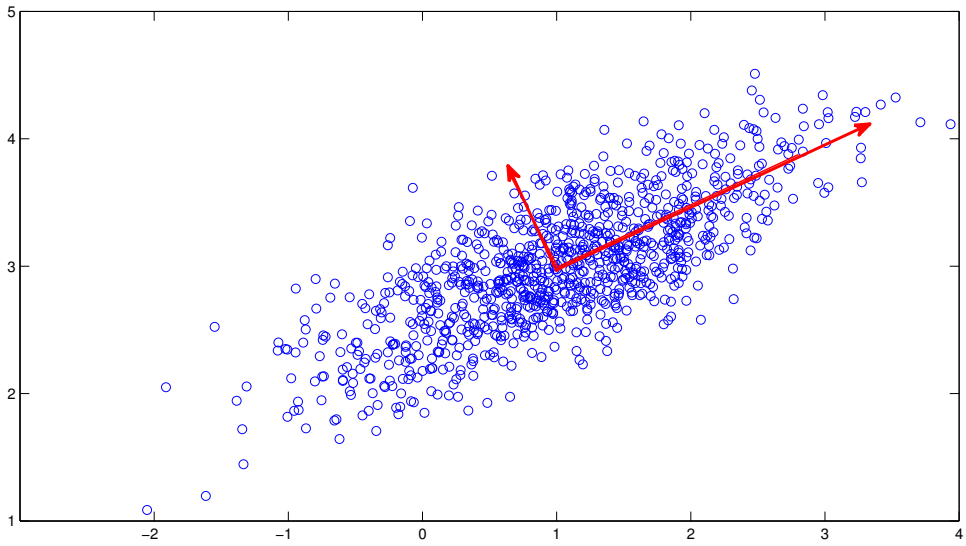


Figure 4.4: Red arrows represent the principal components for this 2-dimensional dataset.

the signal from each axis after another. In other words, if $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$ are column vectors representing each axis after DCT has been applied, $\boldsymbol{x} = [\boldsymbol{d}_1\ \boldsymbol{d}_2\ \boldsymbol{d}_3]^T$.

Now, let us consider PCA more formally. Let $\boldsymbol{x}^i, i = 1, 2, \ldots, N$ be the d-dimensional feature vectors of the training set so that the entire training set can be represented as an $N \times d$ matrix

$$\boldsymbol{X} = \begin{bmatrix} x_1^1 & \ldots & x_d^1 \\ \vdots & \ddots & \vdots \\ x_1^N & \ldots & x_d^N \end{bmatrix}. \tag{4.3}$$

The goal is to map vectors $\boldsymbol{x}^i$ onto vectors $\boldsymbol{z}^i$ in an M-dimensional space $(z_1, \ldots, z_M)$, where $M < d$. Without loss of generality, $\boldsymbol{x}$ can be represented as a linear combination of orthonormal vectors $\boldsymbol{u}_i$

$$\boldsymbol{x} = \sum_{i=1}^{d} z_i \boldsymbol{u}_i \tag{4.4}$$

where the vectors $\boldsymbol{u}_i$ satisfy

$$\boldsymbol{u}_i^T \boldsymbol{u}_j = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{4.5}$$

in which δ_{ij} is known as the Kronecker delta. Each of the coefficients z_i in Eq. (4.4) can be expressed explicitly by using Eq. (4.5) to give

$$z_i = \boldsymbol{u}_i^T \boldsymbol{x}. \tag{4.6}$$

Now, suppose that only a subset $M < d$ of the basis vectors $\boldsymbol{u}_i$ is retained. The remaining $M - d$ coefficients will be replaced by constants b_i so that each vector $\boldsymbol{x}$ is approximated by

$$\tilde{x} = \sum_{i=1}^{M} z_i \boldsymbol{u}_i + \sum_{i=M+1}^{d} b_i \boldsymbol{u}_i. \tag{4.7}$$

Since the original d-dimensional vector $\boldsymbol{x}$ is now approximated by a new vector $\boldsymbol{z}$ with $M < d$ dimensions, this represents a form of dimensionality reduction. The next step is to choose for the basis vectors $\boldsymbol{u}_i$ and coefficients b_i such that the approximation given by Eq. (4.7) gives the best approximation to the original vector $\boldsymbol{x}$ on average for the whole dataset. The error of this dimensionality reduction is thus

$$\boldsymbol{x}^n - \tilde{\boldsymbol{x}}^n = \sum_{i=M+1}^{d} (z_i^n - b_i) \boldsymbol{u}_i. \tag{4.8}$$

Let us define the best approximation to be the one which minimizes the sum of squares of all the errors over the whole dataset. Therefore, we minimize

$$E_M = \frac{1}{2} \sum_{n=1}^{N} \|\boldsymbol{x}^n - \tilde{\boldsymbol{x}}^n\|^2 = \frac{1}{2} \sum_{n=1}^{N} \sum_{i=M+1}^{d} (z_i^n - b_i)^2 \tag{4.9}$$

by utilizing Eq. (4.5). Setting the derivative of E_M with respect to b_i to zero for each i results in

$$\begin{aligned} &\frac{\partial}{\partial b_i} \frac{1}{2} \sum_{n=1}^{N} (z_i^n - b_i)^2 = -\sum_{n=1}^{N} (z_i^n - b_i) = 0 \\ &\Leftrightarrow b_i = \frac{1}{N} \sum_{n=1}^{N} z_i^n = \boldsymbol{u}_i^T \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x} = \boldsymbol{u}_i^T \bar{\boldsymbol{x}}, \end{aligned} \tag{4.10}$$

by using Eq. (4.6) where $\bar{\boldsymbol{x}}$ represents the arithmetic mean of all $\boldsymbol{x}$. Now the error term E_M can be written as

$$\begin{aligned} E_M &= \frac{1}{2} \sum_{n=1}^{N} \sum_{i=M+1}^{d} [\boldsymbol{u}_i^T (\boldsymbol{x}^n - \bar{\boldsymbol{x}})]^2 \\ &= \frac{1}{2} \sum_{i=M+1}^{d} \boldsymbol{u}_i^T \boldsymbol{\Sigma} \boldsymbol{u}_i \end{aligned} \tag{4.11}$$

where $\boldsymbol{\Sigma}$ is commonly known as the covariance matrix of the set of vector $\{\boldsymbol{x}^n\}$ and is defined as

$$\boldsymbol{\Sigma} = \sum_n (\boldsymbol{x}^n - \bar{\boldsymbol{x}})(\boldsymbol{x}^n - \bar{\boldsymbol{x}})^T. \tag{4.12}$$

Theorem 1 (PCA Theorem)**.** *Error term E_M in Eq.* (4.11) *is minimized when the basis vectors are the eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$ defined in Eq.* (4.12)*, i.e.,*

$$\boldsymbol{\Sigma}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i. \tag{4.13}$$

Proof. To minimize Eq. (4.11), let us impose a constraint by requiring the $\boldsymbol{u}_i$ to be orthonormal, which can be taken account using Lagrange multipliers denoted by μ_{ij}. Thus we minimize the function

$$\hat{E}_M = \frac{1}{2}\sum_{i=M+1}^{d} \boldsymbol{u}_i^T \boldsymbol{\Sigma} \boldsymbol{u}_i - \frac{1}{2}\sum_{i=M+1}^{d}\sum_{j=M+1}^{d} \mu_{ij}(\boldsymbol{u}_j^T \boldsymbol{u}_j - \delta_{ij}). \tag{4.14}$$

In matrix notation, this is simplified to

$$\hat{E}_M = \frac{1}{2}\operatorname{Tr}\{\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U}\} - \frac{1}{2}\operatorname{Tr}\{\boldsymbol{M}(\boldsymbol{U}^T\boldsymbol{U} - \boldsymbol{I})\} \tag{4.15}$$

where $\boldsymbol{U}$ is a matrix whose columns represent the eigenvectors $\boldsymbol{u}_i$, $\boldsymbol{M}$ is a matrix with elements μ_{ij} and $\boldsymbol{I}$ is the unit matrix. Now, minimizing Eq. (4.15) with respect to $\boldsymbol{U}$ results in

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{U}}\hat{E}_M = \frac{1}{2}\frac{\partial}{\partial \boldsymbol{U}}\operatorname{Tr}\{\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U}\} - \frac{1}{2}\operatorname{Tr}\{\boldsymbol{M}(\boldsymbol{U}^T\boldsymbol{U} - \boldsymbol{I})\} = 0 \\
&\Leftrightarrow (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}^T)\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{M} + \boldsymbol{M}^T) = 0 \\
&\Leftrightarrow 2\boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{U}2\boldsymbol{M} \\
&\Leftrightarrow \boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{M}
\end{aligned} \tag{4.16}$$

with the help of The Matrix Cookbook (Petersen and Pedersen, 2006, p. 11). Also, the fact that the covariance matrix $\boldsymbol{\Sigma}$ is by definition symmetric and matrix $\boldsymbol{M}$ can be assumed symmetric without loss of generality. Since the matrix $\boldsymbol{U}$ is constructed from eigenvectors $\boldsymbol{u}_i$, it is an orthogonal matrix satisfying $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$. Thus Eq. (4.16) can be equivalently written as

$$\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U} = \boldsymbol{M}. \tag{4.17}$$

Now, let us consider an arbitrary solution of Eq. (4.17) such as

$$\boldsymbol{M}\boldsymbol{\Psi} = \boldsymbol{\Psi}\boldsymbol{\Lambda} \tag{4.18}$$

where $\boldsymbol{\Lambda}$ is a diagononal matrix of eigenvalues. Since $\boldsymbol{M}$ is considered symmetric, the eigenvector matrix $\boldsymbol{\Psi}$ can be chosen to have orthonormal columns, i.e., $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \boldsymbol{I}$. Therefore,

$$\boldsymbol{\Lambda} = \boldsymbol{\Psi}^T\boldsymbol{M}\boldsymbol{\Psi} \tag{4.19}$$

and by substituting Eq. (4.17) into Eq. (4.19), we obtain

$$\begin{aligned} \boldsymbol{\Lambda} &= \boldsymbol{\Psi}^T\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U}\boldsymbol{\Psi} \\ &= (\boldsymbol{U}\boldsymbol{\Psi})^T\boldsymbol{\Sigma}(\boldsymbol{U}\boldsymbol{\Psi}) \\ &= \tilde{\boldsymbol{U}}\Sigma\tilde{\boldsymbol{U}}. \end{aligned} \tag{4.20}$$

Here, we have defined

$$\tilde{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{\Psi}, \tag{4.21}$$

or in other words,

$$\boldsymbol{U} = \tilde{\boldsymbol{U}}\boldsymbol{\Psi}^T. \tag{4.22}$$

This means that an arbitrary solution to Eq. (4.17) can be obtained from the particular solution $\tilde{\boldsymbol{U}}$ by using an orthogonal transformation matrix $\boldsymbol{\Psi}$. It follows that the value of the criterion E_M is invariant under this transformation because

$$\begin{aligned} E_M &= \frac{1}{2}\,\mathrm{Tr}\{\boldsymbol{U}^T\boldsymbol{\Sigma}\boldsymbol{U}\} \\ &= \frac{1}{2}\,\mathrm{Tr}\{\boldsymbol{\Psi}\tilde{\boldsymbol{U}}^T\boldsymbol{\Sigma}\tilde{\boldsymbol{U}}\boldsymbol{\Psi}^T\} \\ &= \frac{1}{2}\,\mathrm{Tr}\{\boldsymbol{\Psi}^T\boldsymbol{\Psi}\tilde{\boldsymbol{U}}^T\boldsymbol{\Sigma}\tilde{\boldsymbol{U}}\} \\ &= \frac{1}{2}\,\mathrm{Tr}\{\tilde{\boldsymbol{U}}^T\boldsymbol{\Sigma}\tilde{\boldsymbol{U}}\} \end{aligned} \tag{4.23}$$

where the invariance property of trace against cyclic permutations is used, as well as the fact that $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \boldsymbol{I}$. Now, since all of the possible solutions give the same value for the error term E_M, it is possible to choose whichever is the most convenient. Thus, the solution given by $\tilde{\boldsymbol{U}}$ is chosen, since this has columns which are the eigenvectors of $\boldsymbol{\Sigma}$, as in Eq. (4.20). (Bishop, 1995) □

In practice, PCA is implemented by centering the feature vectors $\boldsymbol{x}^i$ that construct the data matrix $\boldsymbol{X}$ as in Eq. (4.3), so that $(1/d)\sum_{j=1}^{d} x_j^i = 0$ for $i = 1, 2, \ldots, N$. Then the covariance matrix $\boldsymbol{\Sigma} = (1/N)\boldsymbol{X}^T\boldsymbol{X}$ is a $d \times d$ matrix. The eigenvectors $\boldsymbol{u}_i$ of $\boldsymbol{\Sigma}$

corresponding to the largest eigenvalues are then arranged as the columns of the transform matrix $\boldsymbol{W}$. Now, the Karhunen-Loève transform is defined as

$$\boldsymbol{z} = \boldsymbol{W}^T \boldsymbol{x} \tag{4.24}$$

for an input vector $\boldsymbol{x}$ to be projected to a lower-dimensional space. The dimension of this space is determined by the number of eigenvectors, or principal components, included in the matrix $\boldsymbol{W}$.

4.3 Support Vector Machines

Support vector machines are linear classifiers for a two-class problem where the classes are assumed to be separable in n-dimensional space by a hyperplane. More formally, let $\boldsymbol{x}_i, i = 1, 2, \ldots, N$ be the feature vectors of the training set. Each $\boldsymbol{x}_i$ belong to either of two classes, ω_1, ω_2, which are assumed to be linearly separable. The goal is to design a hyperplane

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0 \tag{4.25}$$

that classifies all the training vectors correctly. Here, $\boldsymbol{w}$ characterizes the hyperplane's direction and w_0 its exact position in space. It is evident by looking at a two-class separation task in Figure 4.5 that a hyperplane satisfying the condition of Eq. (4.25) is not unique. The important question then is, which one of all the possible hyperplanes would be the best choice? For achieving the best generalization of the model, the line in direction 2 in Figure 4.5 is considered the best because it leaves more "room" on either side, so that data in both classes can move a bit more freely, with less risk of causing an error. Thus an SVM classifier in direction 2 (with a margin of $2z_2$) is more trustable than in direction 1 (with a margin of $2z_1$) when operating with unknown data. This is known as the *generalization performance of the classifier*, meaning the capability of the classifier, trained using the training dataset, to operate satisfactorily with data outside this set.

Generally the distance of a point from a hyperplane is given by

$$z = \frac{|g(\boldsymbol{x})|}{\|\boldsymbol{w}\|} \tag{4.26}$$

and parameters $\boldsymbol{w}, w_0$ can be scaled so that the value of $g(\boldsymbol{x})$ at the nearest points in ω_1, ω_2 (Circled in Figure 4.5) is equal to 1 for ω_1 and -1 for ω_2. In practice, this is equivalent to having a margin of $1/\|\boldsymbol{w}\| + 1/\|\boldsymbol{w}\| = 2/\|\boldsymbol{w}\|$ and requiring that

$$\begin{aligned} \boldsymbol{w}^T \boldsymbol{x} + w_0 &\geq 1, \quad \forall \boldsymbol{x} \in \omega_1 \\ \boldsymbol{w}^T \boldsymbol{x} + w_0 &\leq -1, \quad \forall \boldsymbol{x} \in \omega_2. \end{aligned} \tag{4.27}$$

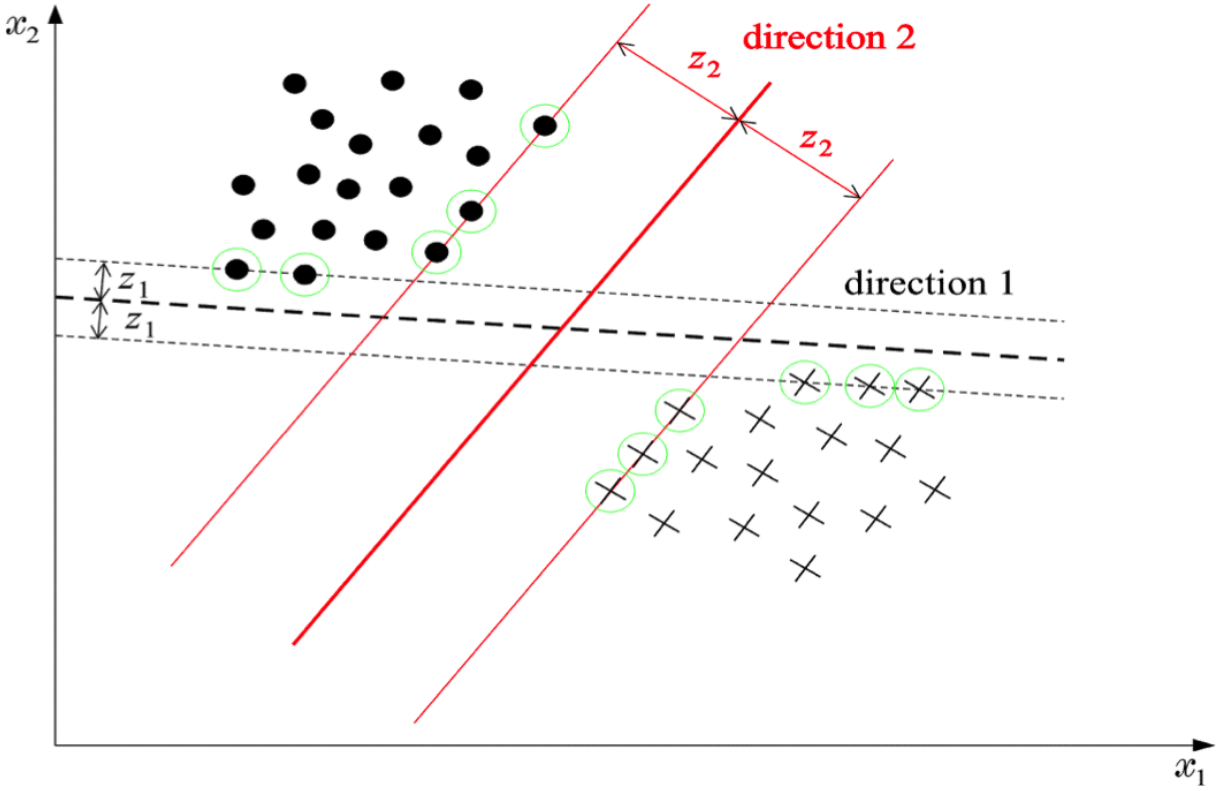


Figure 4.5: The problem of choosing the hyperplane that maximizes the margin between the classes during SVM training phase. (Theodoridis and Koutroumbas, 2006)

Now, let y_i denote the class indicator, i.e., $+1$ for ω_1 and -1 for ω_2. The optimization task can then be formulated as: Compute the parameters $\boldsymbol{w}, w_0$ of the hyperplane so that to:

$$\text{(4.28a)} \qquad \text{minimize} \quad J(\boldsymbol{w}) \equiv \frac{1}{2}\|\boldsymbol{w}\|^2$$

$$\text{(4.28b)} \qquad \text{subject to} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) \geq 1, \quad i = 1, 2, \ldots, N.$$

Clearly, the margin is maximized when the norm is minimized. This is a typical example of a nonlinear quadratic optimization task subject to a set of linear inequality constraints. In order to minimize Eqs. (4.28a, 4.28b), the so-called Karush-Kuhn-Tucker (KKT) con-

ditions have to be satisfied:

$$\frac{\partial}{\partial \boldsymbol{w}}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) = 0 \tag{4.29a}$$

$$\frac{\partial}{\partial w_0}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) = 0 \tag{4.29b}$$

$$\mu_i \geq 0, \quad i = 1, 2, \ldots, N \tag{4.29c}$$

$$\mu_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1] = 0, \quad i = 1, 2, \ldots, N \tag{4.29d}$$

where $\boldsymbol{\mu}$ is the vector of Lagrange multipliers, μ_i, and $\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu})$ is the Lagrangian function

$$\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_{i=1}^{N}\mu_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1]. \tag{4.30}$$

When the condition of Eq. (4.29a) is combined with Eq. (4.30), the direction of the hyperplane can be expressed as

$$\frac{\partial}{\partial \boldsymbol{w}}\mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) = \boldsymbol{w} - \sum_{i=1}^{N}\mu_i[y_i(\boldsymbol{x}_i)]$$

$$\Leftrightarrow \boldsymbol{w} = \sum_{i=1}^{N}\mu_i y_i \boldsymbol{x}_i. \tag{4.31a}$$

Similarly, combining Eq. (4.29b) with Eq. (4.30) results in

$$\sum_{i=1}^{N}\mu_i y_i = 0. \tag{4.32}$$

Since the Lagrange multipliers can be either zero or positive, vector $\boldsymbol{w}$ of the optimal solution is a linear combination of $N_s \leq N$ feature vectors where $\mu_i \neq 0$, that is

$$\boldsymbol{w} = \sum_{i=1}^{N_s}\mu_i y_i \boldsymbol{x}_i. \tag{4.33}$$

These are known as the *support vectors* and thus the optimal hyperplane classifier is known as a support vector machine.

The problem of computing the actual parameters $\boldsymbol{w}$ and w_0, with requirements expressed in Eqs. (4.28a, 4.28b), belongs to *convex programming* family of problems, since the cost function is convex and the set of constraints are linear and define a convex set of

feasible solutions. These kinds of problems can be solved by considering the *Lagrangian duality* so the problem can be stated equivalently in the form

$$\text{maximize} \quad \mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) \tag{4.34a}$$

$$\text{subject to} \quad \boldsymbol{w} = \sum_{i=1}^{N} \mu_i y_i \boldsymbol{x}_i \tag{4.34b}$$

$$\sum_{i=1}^{N} \mu_i y_i = 0 \tag{4.34c}$$

$$\boldsymbol{\mu} \geq \mathbf{0}. \tag{4.34d}$$

By taking the definition of $\boldsymbol{w}$ in Eq. (4.31a) and noting the equivalence in Eq. (4.32), the Lagrangian of Eq. (4.30) can be expressed as

$$\begin{aligned} \mathcal{L}(\boldsymbol{w}, w_0, \boldsymbol{\mu}) &= -\frac{1}{2} \sum_{i=1}^{N} \mu_i y_i \boldsymbol{x}_i^T \sum_{j=1}^{N} \mu_j y_j \boldsymbol{x}_j - \sum_{i=1}^{N} \mu_i y_i w_0 + \sum_{i=1}^{N} \mu_i \\ &= \sum_{i=1}^{N} \mu_i - \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j - w_0 \sum_{i=1}^{N} \mu_i y_i \\ &= \sum_{i=1}^{N} \mu_i - \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \end{aligned} \tag{4.35}$$

and thus the equivalent optimization task becomes

$$\max_{\mu} \left(\sum_{i=1}^{N} \mu_i - \frac{1}{2} \sum_{i,j} \mu_i \mu_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \right) \tag{4.36a}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \mu_i y_i = 0 \tag{4.36b}$$

$$\boldsymbol{\mu} \geq \mathbf{0}. \tag{4.36c}$$

At this point the parameters of the optimization task have changed from $\boldsymbol{w}$ and w_0 into Lagrangian multipliers $\boldsymbol{\mu}$. One important property of Eq. (4.36a) is that the training vectors $\boldsymbol{x}$ are inputted as pairs, representing the inner products. Consequently, the cost function does not depend explicitly on the dimensionality of the input space. Once the optimal $\boldsymbol{\mu}$ has been computed, the optimal hyperplane is obtained via Eq. (4.36a), and w_0 via Eq. (4.29d). (Theodoridis and Koutroumbas, 2006)

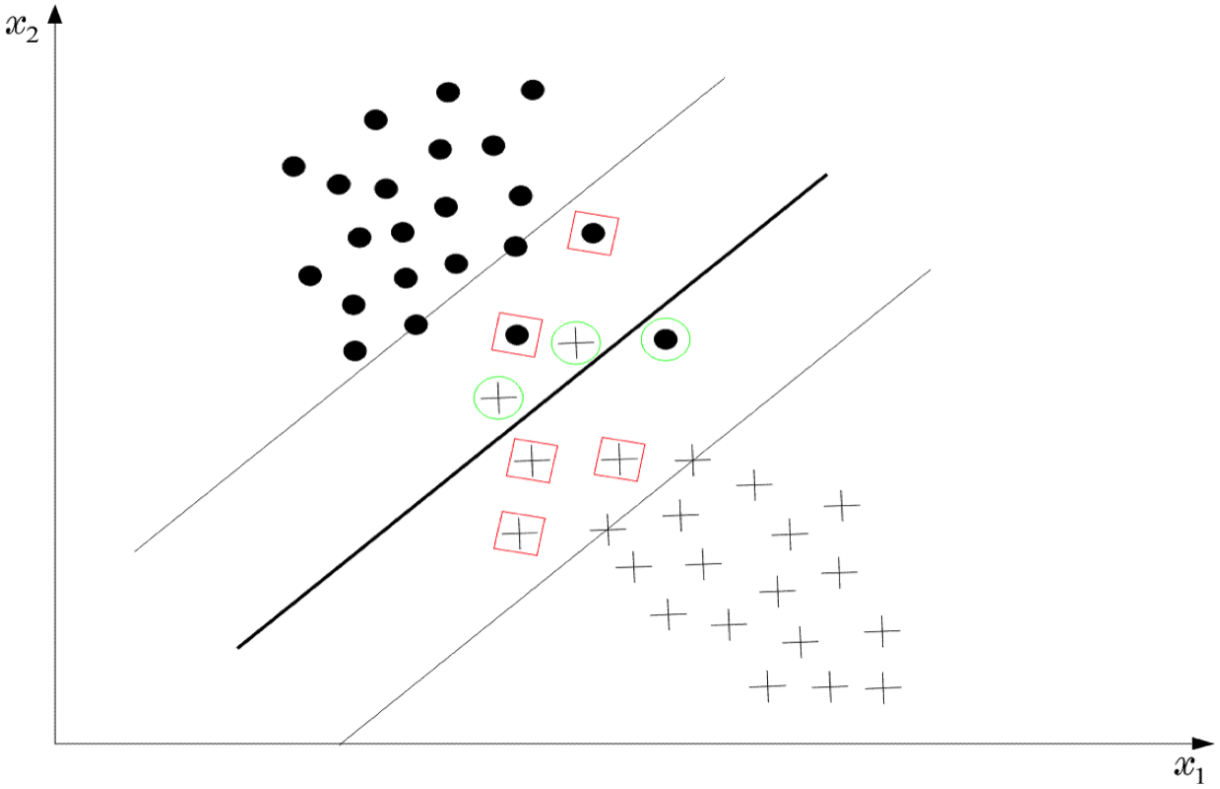


Figure 4.6: Non-separable classes for SVM. Squares around the samples indicate correct classifications while circles indicate misclassifications. The class separation band is indicated by the two thin parallel lines. (Theodoridis and Koutroumbas, 2006)

If the classes are not fully separable by a hyperplane as in Figure 4.6, the classes can still be separated using a class separation band. This means that the vectors falling outside the band are correctly classified and comply with the constraints in Eq. (4.28b), while some of the vectors inside the class separation band are misclassified, causing the nonseparability problem in the first place. In Figure 4.6 the problematic points are enclosed by a circle or a square. This problem can be solved using the *Soft Margin* method proposed by Cortes and Vapnik and is formulated in much more detail in their highly popular article (Cortes and Vapnik, 1995).

Once the SVM classifier has been trained on the training set, the actual classification task of test data requires only applying Eq. (4.25) for each vector in the test set, i.e., a single dot product and a sum of two scalars. Especially when the dimensionality of the data is low, this operation is very fast to compute.

Chapter 5

Experiments

5.1 Artificial Datasets

Three artificial datasets were recorded by three different people. The training set for the DPS model was recorded by the author. Next, two test subjects were recruited to wear the watch for at least a 24-hour period while having 7 to 13 simulated seizures during this time, consisting of trembling in different body and watch orientations with slightly variating periodicity of the trembling.

The first of these test subjects, a female, recorded various day to day activities, including 7 seizures. This dataset was utilized as the validation set, i.e., the parameters and particular activities in the original training set were fine-tuned so that the recognition accuracy on this validation set was maximized. Good performance on validation set would indicate generally good performance. However, if the model has overfitted to the training set, the generalization property of the model may suffer greatly. For this reason, a third dataset was recorded by a third test subject, a male, that was used as the final performance indicator for the model. All the tests were performed on the test set only after the model was already fine-tuned using the validation set.

All training and testing phases were implemented in Matlab 2013b, which had most of the relevant functions inbuilt, such as `dct()`, `svmtrain()` and `eig(cov())`. First step was to preprocess the raw accelerometer signal into feature vectors. In order to do this, the long signal was split into 200 sample windows, each representing 5 s worth of data, with 50% sample overlap between consecutive windows. Next step was to apply DCT on each window for each axis and save the first N DCT components after excluding the highly dominant first DCT component that represents the gravitational acceleration according to He and Jin (2009). This leaves $N \times 3$ components per each window, i.e., $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$ representing the axes. For the PCA, each of these axes are combined into a single $3N$ long column vector: $\boldsymbol{x} = [\boldsymbol{d}_1\ \boldsymbol{d}_2\ \boldsymbol{d}_3]^T$.

To determine a good value for the number of DCT components N, and the number of principal components C, a simple brute force approach was applied. The performance of the model was evaluated on the validation set with N values ranging from 40 to 150 with increases by 5 and C values ranging from 10 to two times the current number of DCT components. From these tests it was determined that most combinations where N was over 45 and C over 40, managed to recognize all seizures correctly, i.e., sensitivity was 100%. From these combinations the one that caused the least number of false alarms, or the highest specificity, would be the best choice. Also, the smaller the parameter values, the better, since that would make the model simpler and prevent overfitting. In Figure 5.1 the amount of false positives is visualised and based on these experiments, values $N = 55, C = 40$ were chosen as the parameters for running the DPS model on the test set. Model complexity is increasing from the upper left corner to the bottom right corner in Figure 5.1, so parameter combinations near upper left corner are assumed to be better. In practice, this means that during the 42 h period, all 7 seizures were recognized correctly along with 29 false alarms occuring during 16 different events. Based on the diary of the test subject, these false positives occurred most likely during brushing teeth, washing hands, shaking a bottle, stroking dogs, and while cleaning up. Especially washing hands was an activity that was easily added to the training set, after obtaining the validation set and asking the test subject of possible activities during the false alarms. It was clearly seen that including this 1 minute dataset of washing hands actually reduced the number of false alarms, even when these activities were performed by two different people and not knowing the exact movements the hands of the second test subject would make while washing them.

Using the same model for the test set, recorded by the third test subject, resulted in recognizing 12 seizures out of 13 while causing 6 unique false alarms of which 3 occurred during brushing teeth (around 5 minutes in total causing multiple alarms continuously during the same activity). From the rest, 2 occured within 2 minutes of brushing teeth and 1 false alarm (5 s) remains unexplained. All results are listed in Table 5.1 and because there is no information on how many seizure-like events the test subjects had, specificity is not included. Of course there were a total of 35342 time windows in the test set of which only around 200 were during a seizure and 82 were wrongly labeled as seizures. However, this is mostly due to particular test subjects daily activities not including so many kinds of seizure-like activities.

5.2 Epilepsy Datasets

Early on in this project it was decided that since the epilepsy datasets were not labeled well enough nor is there any video material of the seizures recorded, there was a real

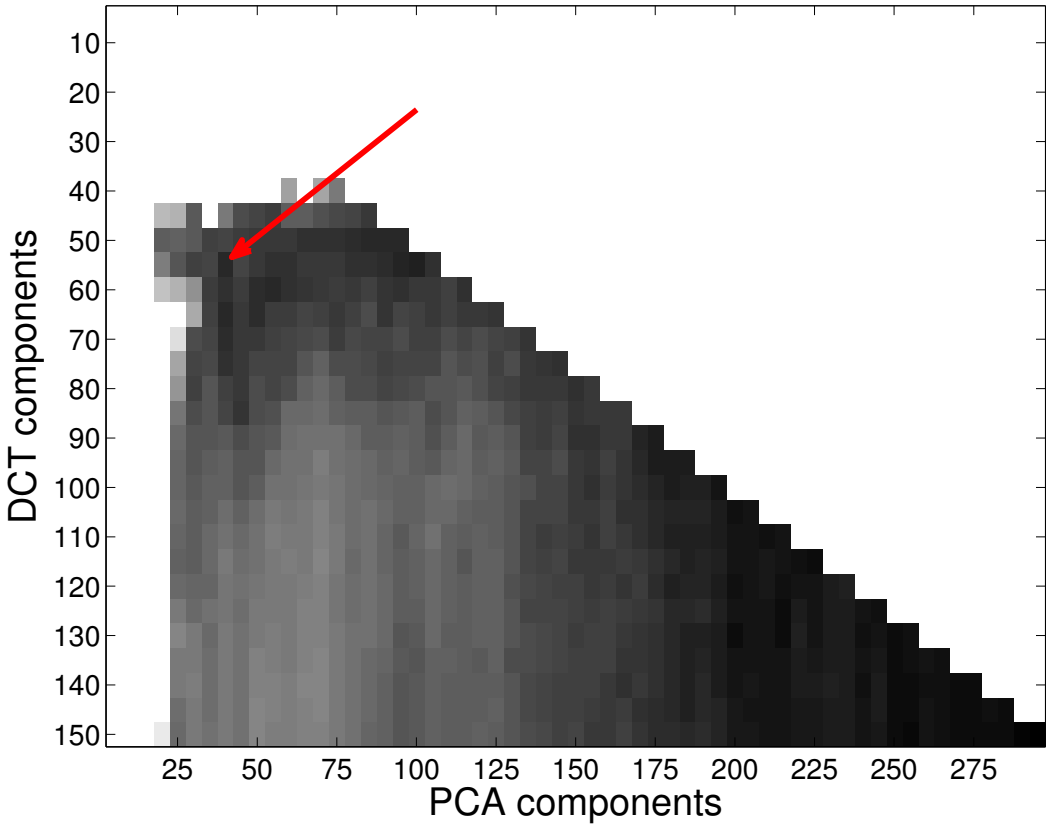


Figure 5.1: The number of false positives are displayed as a function of the number of DCT (N) and PCA (C) components. Whiter values mean greater number of false alarms or SVM not being able to converge while darker values indicate better performance of the classifier. Red arrow points to the (N, C) combination chosen.

need for artificial datasets. However, when applying the same DPS model to the original epilepsy datasets, the results were quite ambiguous. Table 3.1 shows how the length of the observations labeled by nurses differs from that of data, as mentioned in Section 3.3. The actual timestamps for the estimated seizures mostly do not match at all, and it is problematic to determine how much difference there is between particular points of time since the labeled length and data length of seizures do not match. However, when considering an arbitrary model trying to recognize certain events from time series, it would be reasonable to assume that either the event is very rare or difficult to recognize, thus resulting in zero events. Or, the event would be something very easily recognizable

Table 5.1: Results on validation and test set.

	Validation Set	Test Set
Number of seizures (s)	7	13
Seizures recognized (r)	7	12
Total false positive events	29	82
Unique false positives	16	6
Sensitivity (s/r)	100%	92.3%

so that hundreds of events would be recognized. With the real world epilepsy datasets, out of several days of ACM data, only 1–12 events are recognized per dataset whereas nurses have labeled 1–7 seizures per dataset (see Table 5.2). Also, the number of seizures labeled is only the lower limit to the actual number of seizures which occured during the monitoring phase, since there were neither cameras nor nurses around 24/7. Thus, the number of predicted seizures being, as one might say, "in the same ballpark" than the actual number of seizures does indicate that the DPS model recognizes something that might be relevant to an epileptic seizure. Table 5.2 shows that some of the detected events are very long, even over a minute long. Since the patients wore the watch on their non-dominant hand, there remains the question of what else could this movement have been if not a seizure. Especially since we know for a fact that there were some seizures during the monitoring phase. This means that it is likely that at least some of the events detected are actually real epileptic seizures, though this claim cannot be confirmed from these datasets only. Interestingly, since there was no video material associated with the epilepsy datasets, it was not actually known *exactly* what kind of movement the model should even be recognizing. However, with a high probability it was very similar to the movements in the artificial seizure training set, which was based on epilepsy material provided by Vivago (Epilepsy Society, 2013).

Table 5.2: Properties and results on real life epilepsy dataset.

Dataset	Seizures labeled	Unique events detected	Longest seizure detected
L1002_04	7	7	80 s
L1002_13	4	10	15 s
L1002_28	1	12	108 s
L1004_16	1	4	8 s
L1008_21	4	1	35 s
L1011_16	3	6	5 s

Chapter 6

Discussion and Conclusion

In this thesis, the problem of detecting motor seizures during epileptic attacks using triaxial ACM data has been approached by using a DCT–PCA–SVM (DPS) model that first preprocesses each axis using DCT. Then the DPS model combines the data and lowers its dimension by PCA, and finally classifies the time window into a seizure or nonseizure by SVM. Multiple studies and their results in epileptic seizure detection and human activity recognition were presented. Different types of epileptic seizures were detailed as well as the artificial and real life ACM datasets.

The DPS model introduced by He and Jin (2009) was explained through a diagram and its parts were detailed in a mathematically sound way. Discrete cosine transform and principal component analysis were defined and a proof of the PCA theorem was presented. The theory behind support vector machines was presented in a case where the classes are separable.

For the experiments, artificial datasets were recorded by three different test subjects so that one was used for training of the DPS model, second to validate the model and fine-tune its training parameters, and the third one was used for measuring the performance of the model on previously unseen data. The results on the six epilepsy datasets were also presented, but since there is no exact information of what to look for, the results are highly ambiguous.

The answer to the original research question proposed in Chapter 1 is that:

> *Based on the artificial ACM datasets, recorded with a single wrist-worn triaxial accelerometer, around 92% of the motor seizures were detected by using the DPS model and resulting to a reasonably low number of false alarms.*

The performance of the DPS model would likely be further increased by using better training data. It cannot be stressed enough that:

The experimental setting in this thesis was set to be more difficult than the real world case, due to test subjects wearing the sensor on their dominant wrist, where the potential of recognizing false seizure-like movements was much greater than during de facto convention of using non-dominant wrist.

The reason for this decision was to better analyze the false alarms caused by everyday activities where most seizure-like activities are performed using the dominant hand, such as brushing teeth or playing guitar. Thus the artificial datasets reveal more information on false alarms, but in a real world case where an epileptic seizure affects the whole body symmetrically, wearing the sensor on the non-dominant wrist would not affect the recognition performance of the model, but it would definitely lower the number of false alarms. The experiments showed that epileptic seizures were quite easily recognized and only 1 out of 13 seizures was not recognized in the test set. Most of the false alarms related to brushing teeth or washing hands. Again, if the test subjects had worn the sensor in their non-dominant hand, brushing teeth would not have caused a false alarm. In conclusion, the DPS model was shown to work very well in a setting where the interesting movements were defined well enough, even if they were performed by different people.

While the results on the real epilepsy datasets are ambiguous, there is some support to the idea that some of the alarms detected by DPS were in fact real seizures that occured in the datasets due to the low number of alarms and their length. One could also argue that the simple fact that all the artificial data was recorded while wearing the sensor in the right arm, and 5 out of 6 epilepsy datasets being recorded while wearing the sensor in the left arm, might make this more difficult. This problem was not studied at all in this thesis, but based on how the model makes orientation of the sensor practically irrelevant, this should not be the primary cause of ambiguous results. Furthermore, He and Jin (2009) did not fix the sensor on the test subjects' bodies: instead they placed the sensor in the subjects' trouser pockets to achieve robustness in regards to sensor position.

Finally, the DPS algorithm implemented in this thesis has potential for much more ambitious scenarios in human activity recognition in general as demonstrated in He and Jin (2009). The DPS algorithm is not particularly "tailor-made" for epilepsy detection, rather, it should work in many different kinds of human activity recognition tasks. It is unknown how well the DPS algorithm would perform for very short events, such as 1 s (40 Hz) long. It is probable that due to DCT, the sample window should be at least couple of seconds long, or the sampling rate should be higher. For example in (He and Jin, 2009), 100 Hz and 5.12 s long sample window was used. In some sense the DPS model could be considered a brute force approach to epileptic seizure detection, but in order to accurately capture the periodicity and the multitude of limb movements involved in a seizure, some FFT/DCT type technique should be utilized anyway. Simple models, like the one proposed by Cuppens et al. (2009), may work well enough for epileptic seizure

detection in certain environments, such as during the night with an accelerometer attached to each limb. However, for a real world case including a multitude of movements, but only one accelerometer, a very general model like DPS is needed for seizure detection.

6.1 Future Work

The DPS model could quite easily be extended to using multiple sensors, or from a mathematical point of view, singular axes. More formally, the input to the PCA in the scope of this thesis is defined as $\boldsymbol{x} = [\boldsymbol{d}_1\ \boldsymbol{d}_2\ \boldsymbol{d}_3]^T$, where $\boldsymbol{d}_i$ represents a single axis of the sensor. However, there is no limitation on the number of sensors. thus the input might as well be $\boldsymbol{x'} = [\boldsymbol{d}_1\ \boldsymbol{d}_2\ \ldots \boldsymbol{d}_n]^T$ when using n acceleration axes. In theory, since the DCT components already contain most of the information, adding more sensors and probably taking more PCA components for SVM, would only assist in distinguishing between different events. This would then increase the performance of the classifier. Important requirement for these new axes would be that the position would not change during the training and testing phase, but this is obviously a requirement for all the experiments performed in this thesis.

Another idea is related to the DCT feature extraction heuristic proposed by He and Jin (2009). Due to the energy-compaction property of DCT, they simply choose the first n DCT components while disregarding the very first one. It might be the case that actually disregarding the first 20 or 30 components would capture more information of the signal, since most of the information seems to be contained by the later components (see Figure 4.3d).

It is likely that this research will be continued in collaboration with Vivago and HUS, and hopefully there will be a possibility to apply the DPS model to a larger and better labeled epilepsy dataset. This would most likely enable better training of the model and definitely enable more conclusive testing of its performance. It is especially interesting to consider how well the model could generalize to different users, or would it be a better approach to train the model more individually to each user, such that its seizure classification performance would increase with more usage.

Bibliography

Ahmed, N., Natarajan, T., and Rao, K. R. (1974). Discrete cosine transform. *Computers, IEEE Transactions on*, 100(1):90–93.

Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer.

Begley, C. E., Famulari, M., Annegers, J. F., Lairson, D. R., Reynolds, T. F., Coan, S., Dubinsky, S., Newmark, M. E., Leibson, C., So, E., et al. (2000). The cost of epilepsy in the united states: An estimate from population-based clinical and survey data. *Epilepsia*, 41(3):342–351.

Beniczky, S., Polster, T., Kjaer, T. W., and Hjalgrim, H. (2013). Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: A prospective, multicenter study. *Epilepsia*, pages n/a–n/a.

Bersch, S., Chislett, C. M. J., Azzi, D., Khusainov, R., and Briggs, J. (2011). Activity detection using frequency analysis and off-the-shelf devices: Fall detection from accelerometer data. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pages 362–365.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

Bouten, C., Koekkoek, K., Verduin, M., Kodde, R., and Janssen, J. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *Biomedical Engineering, IEEE Transactions on*, 44(3):136–147.

Cios, K., Pedrycz, W., and Swiniarski, R. (1998). *Data Mining Methods for Knowledge Discovery*. Kluwer Academic Publishers.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Cuppens, K., Lagae, L., Ceulemans, B., Van Huffel, S., and Vanrumste, B. (2009). Detection of nocturnal frontal lobe seizures in pediatric patients by means of accelerometers: A first study. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 6608–6611.

Deckers, C., Genton, P., Sills, G., and Schmidt, D. (2003). Current limitations of antiepileptic drug therapy: a conference review. *Epilepsy research*, 53(1):1–17.

Degen, T., Jaeckel, H., Rufer, M., and Wyss, S. (2003). Speedy: a fall detector in a wrist watch. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, page 184. IEEE Computer Society.

Doughty, K., Lewis, R., and McIntosh, A. (2000). The design of a practical and reliable fall detector for community and institutional telecare. *Journal of Telemedicine and Telecare*, 6(suppl 1):150–154.

Elevant, J. (1999). *Monitoring Epilepsy With a Wrist Carried Motion Sensor*. PhD thesis, M. Sc. thesis, Royal Institute of Technology, Stockholm.

Epilepsy Society (2013). Epilepsy seizures. `http://www.epilepsysociety.org.uk/aboutepilepsy/whatisepilepsy/seizures`. Accessed June 8, 2013.

Förster, K., Roggen, D., and Troster, G. (2009). Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain? In *Wearable Computers, 2009. ISWC '09. International Symposium on*, pages 77–84.

He, Z. and Jin, L. (2009). Activity recognition from acceleration data based on discrete consine transform and svm. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 5041–5044. IEEE.

He, Z.-Y. and Jin, L.-W. (2008). Activity recognition from acceleration data using ar model representation and svm. In *Machine Learning and Cybernetics, 2008 International Conference on*, volume 4, pages 2245–2250.

Jallon, P. (2010). A bayesian approach for epileptic seizures detection with 3d accelerometers sensors. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 6325–6328. IEEE.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, second edition.

Kangas, M., Konttila, A., Lindgren, P., Winblad, I., and Jämsä, T. (2008). Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait & posture*, 28(2):285–291.

Khan, A., Lee, Y.-K., Lee, S., and Kim, T.-S. (2010). A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *Information Technology in Biomedicine, IEEE Transactions on*, 14(5):1166–1172.

Lee, S., Le, H. X., Ngo, H. Q., Kim, H. I., Han, M., Lee, Y.-K., et al. (2011). Semi-markov conditional random fields for accelerometer-based activity recognition. *Applied Intelligence*, 35(2):226–241.

Lockman, J., Fisher, R. S., and Olson, D. M. (2011). Detection of seizure-like movements using a wrist accelerometer. *Epilepsy & Behavior*, 20(4):638–641.

Maurer, U., Smailagic, A., Siewiorek, D. P., and Deisher, M. (2006). Activity recognition and monitoring using multiple sensors on different body positions. In *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, pages 4–pp. IEEE.

Metsähonkala, E.-L. (2012). Epilepsy dataset recorded at HUS, The Hospital District of Helsinki and Uusimaa.

Microsoft (2013). Kinect. `http://www.xbox.com/en-US/kinect`. Accessed June 8, 2013.

Moiz, F., Natoo, P., Derakhshani, R., and Leon-Salas, W. (2011). A comparative study of classification methods for gesture recognition using a 3-axis accelerometer. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2479–2486.

Nijsen, T., Aarts, R., Cluitmans, P. J. M., and Griep, P. (2010). Time-frequency analysis of accelerometry data for detection of myoclonic seizures. *Information Technology in Biomedicine, IEEE Transactions on*, 14(5):1197–1203.

Nijsen, T., Arends, J., Griep, P., and Cluitmans, P. (2005). The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy. *Epilepsy & behavior: E&B*, 7(1):74.

Nijsen, T. M., Cluitmans, P. J., Griep, P. A., and Aarts, R. M. (2006). Short time fourier and wavelet transform for accelerometric detection of myoclonic seizures. *IEEE/EMBS Benelux*, pages 7–8.

O'Brien, J. W. (2005). The jpeg image compression algorithm.

Panayiotopoulos, C. P. (2006). A practical guide to childhood epilepsies - video-EEGs. [CD-ROM].

Petersen, K. B. and Pedersen, M. S. (2006). The matrix cookbook.

Pitkänen, A., Schwartzkroin, P., and Moshé, S. (2005). *Models of Seizures and Epilepsy*. Elsevier Science.

Ravi, N., Dandekar, N., Mysore, P., and Littman, M. L. (2005). Activity recognition from accelerometer data. In *Proceedings of the national conference on artificial intelligence*, volume 20, page 1541. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Schulc, E., Unterberger, I., Saboor, S., Hilbe, J., Ertl, M., Ammenwerth, E., Trinka, E., and Them, C. (2011). Measurement and quantification of generalized tonic–clonic seizures in epilepsy patients by means of accelerometry—an explorative study. *Epilepsy research*, 95(1):173–183.

Schulze-Bonhage, A., Sales, F., Wagner, K., Teotonio, R., Carius, A., Schelle, A., and Ihle, M. (2010). Views of patients with epilepsy on seizure prediction devices. *Epilepsy & behavior*, 18(4):388–396.

Stikic, M. and Schiele, B. (2009). Activity recognition from sparsely labeled data using multi-instance learning. *Location and Context Awareness*, pages 156–173.

Struska, J. (2013). Gtc/grand mal seizure. `http://www.youtube.com/watch?v=Nds2U4CzvC4`. Accessed October 3, 2013.

Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*. Academic Press, 3rd edition.

Tuominen, P. (2013). Diskreetin aallokemuutoksen käyttö signaalin pakkauksessa. Bachelor's thesis, University of Helsinki.

Vail, D. L., Veloso, M. M., and Lafferty, J. D. (2007). Conditional random fields for activity recognition. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, page 235. ACM.

Yang, J.-Y., Wang, J.-S., and Chen, Y.-P. (2008). Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern recognition letters*, 29(16):2213–2220.