



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

PROGRAMA DE DOCTORAT EN ENGINYERIA BIOMÈDICA
DEPARTAMENT D'ENGINYERIA DE SISTEMES,
AUTOMÀTICA I INFORMÀTICA INDUSTRIAL
CENTRE DE RECERCA EN ENGINYERIA BIOMÈDICA

Tesis Doctoral por compendio de publicaciones

Muscular pattern based on multichannel surface EMG during voluntary contractions of the upper-limb

Mislav Jordanic

September 2017

Directores:

Miguel Angel Mañanas Villanueva

Mónica Rojas-Martínez

Abstract

Text here

Resumen

Texto aqui!

Contents

Abstract	i
Resumen	ii
List of Tables	v
List of Figures	vii
1 Introduction	1
1.1 Muscle anatomy	1
1.2 Surface electromyography	3
1.3 Task identification	4
1.3.1 Pattern recognition	6
Linear Discriminant Analysis	7
Support Vector Machine	11
Bibliography	18

List of Tables

List of Figures

1.1	Four types of recording surface EMG signal: monopolar, bipolar, linear electrode array, HD-EMG. Figure was modified from (Merletti et al., 2010) Traditional monopolar detection with respect to a remote reference taken as zero (reference) potential. b) Bipolar (or single differential, SD) detection (or montage) along the fiber direction. c) Linear (one dimensional, 1-D) array of electrodes along the fiber direction. Spatial filters (such as double or N-differential) can be obtained by properly weighting and adding the signals from nearby electrodes. d) 2-dimensional (2-D) array of electrodes providing an image of potential distribution. Spatial filters (such as double or N-differential in the column or row direction, Laplacian, inverse binomial, etc.) may be applied to the signals. Image processing procedures may be applied to the “image” to interpolate or virtually rotate the image to align it with the fiber direction or to detect edges or areas of high or low activity or gradient. are	4
1.2	The figure represents the HD-EMG activation map recorded on the biceps brachii muscle during flexion. Distinct activation of the two heads can be noticed in the map. Modified from Monica	5

Chapter 1

Introduction

1.1 Muscle anatomy

Movement is achieved using over 600 skeletal muscles, the ones that can be controlled voluntarily.

Smooth muscles - mostly in internal organs

cardiac muscle

Elementary building block of a muscle is muscle cell, or muscle fiber - *myocyte*. Myocytes are ensheathed by *endomysium*, a connective tissue that contains nerves and capillaries. They are organized in bundles of 10 to 100, *fascicles*, and surrounded by sheath of connective tissue, *perymysium*. Group of fascicles is finally grouped together and enveloped by *epimysium*, forming a muscle.

Sarcolema is the cell membrane of myocyte, consisting of a lipid bilayer that contains intracellular liquid, *myoplasma*. In the myoplasma, thin and thick filaments are serially connected, forming *sarcomeres*, which are longitudinally connected in *myofibrils* that extend through entire length of the myocyte. During shortening of muscle fibers, thin and thick filaments of sarcomeres are pulled together by cross-bridges between them. Total shortening of myofibril is summation of shortenings of sarcomeres of which it is composed.

Each motor neuron at the neuromuscular junction innervates several muscle fibers, forming the

smallest functional unit called *motor unit*. It was firstly defined by Liddell and Sherrington in 1925 (Liddell and Sherrington, 1925; Sherrington, 1925) and is composed of motor neuron with axon and dendrites, and muscle fibers that axon innervates (Duchateau and Enoka, 2011). Since motor neuron with a single action potential usually evokes action potentials simultaneously in all belonging muscle fibers, by observing action potentials of the muscle fibers, information on activity of motor neurons in spinal cord or brain stem can be inferred (Merletti and Farina, 2016).

Motor neurons transfer nerve impulses that control the muscle from spinal cord to neuromuscular junction. At the nerve endings action potentials trigger the release of the neurotransmitter *acetylcholine* to the narrow space between the axon and sarcolemma, which causes sodium channels in sarcolemma to open. If the amount of acetylcholine is sufficient for the excitation, depolarization takes place and the action potential propagates longitudinally from the junction towards the ends of the fiber causing contraction (Henneberg, 1999). Speed of action potential propagation is called *conduction velocity* and typically ranges around 4 m/s.

Detailed analysis of muscle physiology can be found elsewhere (Squire, 1986).

Pool of motor neurons that innervates entire muscle generally ranges from ten to thousand, depending on the muscle. (Merletti and Farina, 2016).

Monice

By the characteristics of muscle fiber, there are three main types of muscle fibers:

- Fast twitch, fatigable fibers (FF, or type IIb)
They have high levels of ATP for anaerobic energy supply. They are present in pale muscles. glycolytic and work well in ischemic or low oxygen conditions. fast twitch, large forces. high nerve conduction velocity.
- Fast twitch, fatigue-resistant (FR, or type IIa) oxidative glycolytic, have fast twitch and are resistant to fatigue. intermediate conduction velocity
- Slow twitch, very resistant to fatigue (S, or type I) Oxidative and do not work well in low oxygen conditions. They generate small forces, have slow twitch and resistant to fatigue. lower nerve conduction velocity. This fiber type is very resilient to fatigue because of high

oxidative metabolism and energy efficiency. They are present in high percentage in red muscles, such as soleus.

Muscle fibers innervated by the same motor neuron have similar histochemical and contractile characteristics.

Force that muscle fibers generate depends on firing frequency of the action potentials (rate coding) innervating the neuromuscular junction, and the recruitment strategy by which the motor units are activated, i.e., the number of activated motor units. Firing frequency and the recruitment strategy depend on the speed and force of contraction. Muscle units with low threshold are activated firstly, resulting in low force and high endurance, i.e., resistance to fatigue. If greater force is required, muscle units with higher threshold that are prone to fatigue are activated (Freund et al., 1975; Merletti and Parker, 2004). This was first proposed by Henneman et al. in 1965 (Henneman et al., 1965), who state that order of recruitment of motor neurons is based on size principle, that is, neurons with smaller axons are recruited at lower effort levels and with increase in force, larger motoneurons are recruited. Therefore, S type muscle units, which have the smallest motoneurons are recruited first, followed by FR type units, and finally FF units.

1.2 Surface electromyography

Myoelectric signal is a superposition of electrical activity (propagating action potentials) produced by the muscle fibers while contracting. It can be characterized (Farina et al., 2014)

EMG signals could be recorded either non-invasively (surface EMG, sEMG) or invasively with needle and wire electrodes (intramuscular EMG, iEMG) (Marateb et al., 1999). Although iEMG signal is usually with higher quality (in terms of signal-to-noise ratio), it was shown that both approaches provide a similar quality of identification of upper-arm motor task (Hargrove et al., 2007).

Depending on number of electrodes used for the recording. the following classification exists: monopolar, bipolar, linear electrode array, and high-density EMG (see Figure 1.1)

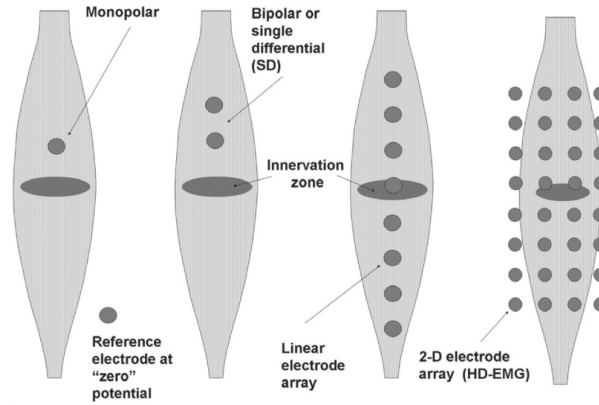


Figure 1.1: Four types of recording surface EMG signal: monopolar, bipolar, linear electrode array, HD-EMG. Figure was modified from (Merletti et al., 2010) Traditional monopolar detection with respect to a remote reference taken as zero (reference) potential. b) Bipolar (or single differential, SD) detection (or montage) along the fiber direction. c) Linear (one dimensional, 1-D) array of electrodes along the fiber direction. Spatial filters (such as double or N-differential) can be obtained by properly weighting and adding the signals from nearby electrodes. d) 2-dimensional (2-D) array of electrodes providing an image of potential distribution. Spatial filters (such as double or N-differential in the column or row direction, Laplacian, inverse binomial, etc.) may be applied to the signals. Image processing procedures may be applied to the “image” to interpolate or virtually rotate the image to align it with the fiber direction or to detect edges or areas of high or low activity or gradient. are

HD-EMG 1.2

1.3 Task identification

The central nervous system (CNS) is responsible for processing information received from all parts of the body. The two main organs of the CNS are the brain and the spinal cord and are entirely composed of two kinds of specialized cells: neurons and glia. The brain is the most complex part of the human body and exerts a centralized control over the other organs. Neurons, the basic working units of the brain, are designed to transmit information within the brain to other nerve cells and to communicate with muscles and gland cells. The complex architecture of the brain is built on the extensive number of interconnected neurons sharing information through specialized connections called synapses. This connection allows neurons to communicate through an electrical or chemical signals, producing ionic currents that generate electric and magnetic fields.

The CNS is organized in multiple levels, from simple connections between cells to coordinated

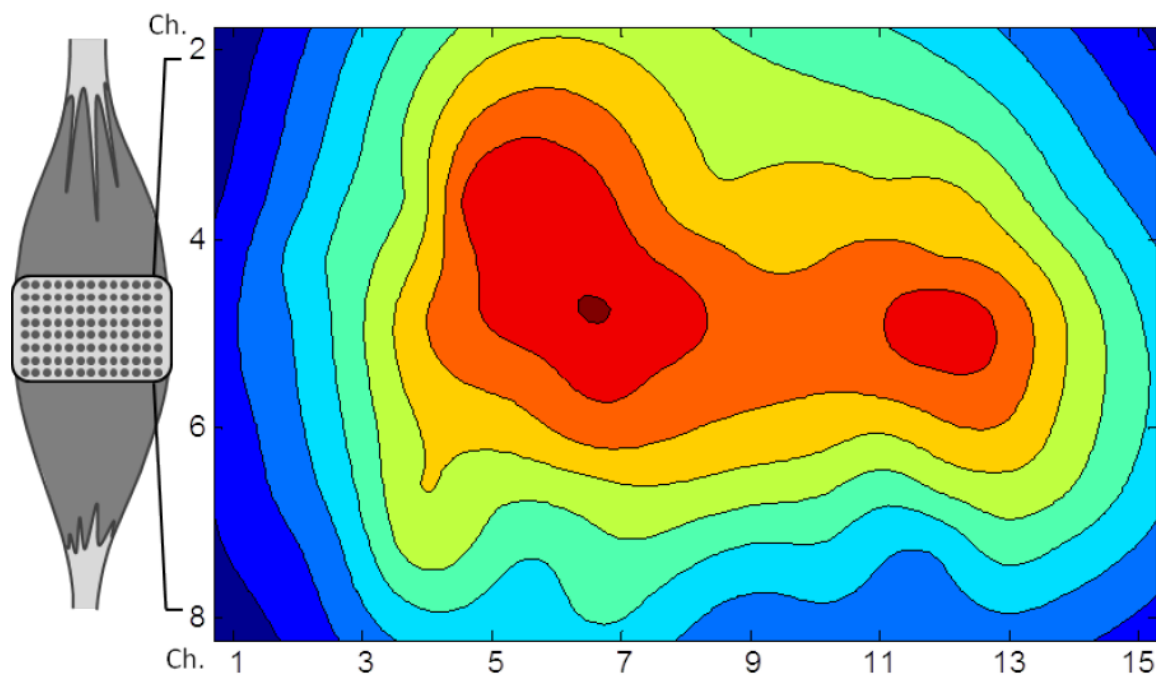


Figure 1.2: The figure represents the HD-EMG activation map recorded on the biceps brachii muscle during flexion. Distinct activation of the two heads can be noticed in the map. Modified from Monica

cell populations, building a complex architecture of interconnected brain regions. The neural processes at this last level are produced by the dynamic coordination of smaller elements. In the cerebral cortex, all this brain activity is summed and its electric and magnetic fields can be measured on the scalp surface.

Pattern recognition is an alternative to conventional control algorithms. The prerequisite of using pattern recognition for task identification is the presence of a pattern that can be extracted from the EMG signal. Major advancement over conventional conventional switching myocontrol is the possibility of.

Pattern recognition approach doesn't support proportional and simultaneous control for multiple motor tasks. Therefore, tasks need to be performed sequentially. This type of control prevents the user from achieving a fluid movement, but also demands planning of movement execution. Although pattern recognition improves the possibility, it has serious limitations.

This implies sequential control which prevents the subject from doing fluent, e.g. Davidge et al. designed a system where movements that combine DoFs are labeled as unique classes in LDA problem, whereas Young et al. (Young et al., 2013) propose system of parallel LDA classifiers

that use conditional probabilities to separate between combination of tasks.

Proportional review (Fougner et al., 2012) Force can be estimated based on the EMG : (Staudenmann et al., 2010)

In pattern recognition, there is still a large gap between industry and practice (Jiang et al., 2012).

On the other hand, one of the disadvantages of pattern recognition is the fact that in spite of the high accuracy, an error could lead to the completely unwanted task. Also, although identification rate is usually very high during the stationary task, errors often occur during transition between tasks. This problems can be partially prevented by employing the e.g. majority voting principle (Englehart and Hudgins, 2003) (300ms, LDA), or decision-based velocity ramp that attenuates the velocity of a movement after the change of a task (Simon et al., 2011).

Challenges in pattern recognition are electrode shift (Hargrove et al., 2008; Young et al., 2011), change in arm posture (Fougner et al., 2011), slow time dependent changes (Farina et al., 2014) such as fatigue (Tkach et al., 2010), and change in electrode-skin impedance (Clancy et al., 2002).

Future works: dynamic system, hybrid system

1.3.1 Pattern recognition

Pattern recognition is a classification technique and the principle by which it is performed is learned independently from the data, i.e., training set. There are two main types of pattern recognition: supervised and unsupervised. Supervised pattern recognition implies that the classes of the training set are known and are used to obtain the model. New inputs are identified as one of the predetermined classes. On the other hand, unsupervised pattern recognition is used when no labels are available and samples are assigned to unknown classes. This technique is more appropriate for the clustering problem because the classes are determined automatically by the system, whereas supervised approach is more appropriate for the classification because classes are defined by the system designer, and, therefore, it is usually used in task identification based on EMG.

In statistical pattern recognition, each sample is composed of m measures that form the pattern, i.e., features $(x_0, x_1, \dots, x_{m-1})$. The objective of the algorithm is to obtain a decision rule, i.e., the decision boundary which separates well samples of different classes. There are many *state-of-the-art* classifiers that use various principles to construct these boundaries. However, many researchers agree that the fidelity of the classification in EMG applications depends mostly on selection of features. In other words, with appropriate selection of features, all classifiers will give similar classification result. A short introduction is provided on the two methods used in the thesis: Linear Discriminant Analysis and Support Vector Machine.

Linear Discriminant Analysis

All models are wrong; some models are useful.

George E. P. Box

Linear Discriminant Analysis is a computationally simple and efficient classifier with linear decision boundary and it is based on the Bayesian equation. In a classical problem with n samples in training set, which consist of m features, the dataset of available samples is a matrix of dimension $[n \times m]$, whereas the label that describes the belonging of each sample to one of the classes is y , where $y \in (0, 1, 2, \dots, K - 1)$.

According to Bayesian equation, the probability that a sample \mathbf{x}_0 belongs to a class k is equivalent to the:

$$P(y = k | \mathbf{x} = \mathbf{x}_0) = \frac{P(\mathbf{x} = \mathbf{x}_0 | y = k) P(y = k)}{P(\mathbf{x} = \mathbf{x}_0)} \quad (1.1)$$

, where k represents the class. Term $P(\mathbf{x} = \mathbf{x}_0 | y = k)$ is called the *class-conditional* probability and describes the probability that the sample with exact features \mathbf{x}_0 is encountered within the group of samples belonging to the class k . Term $P(y = k)$ is called the *a priori* probability and describes the probability that the sample belonging to the class k is found within the group of all samples, regardless of the features. Finally, the term $P(\mathbf{x} = \mathbf{x}_0)$ is called the *marginal* probability and describes the probability of finding the sample with exact set of features in the dataset, regardless of the class. Marginal probability can be written as a sum of class-conditional

probabilities multiplied by the a priori probabilities for each class:

$$\begin{aligned}
 P(\mathbf{x} = \mathbf{x}_0) &= P(\mathbf{x} = \mathbf{x}_0 \mid y = 1) P(y = 1) + \\
 &P(\mathbf{x} = \mathbf{x}_0 \mid y = 2) P(y = 2) + \cdots + \\
 &P(\mathbf{x} = \mathbf{x}_0 \mid y = K) P(y = K)
 \end{aligned} \tag{1.2}$$

Following the Bayesian theory, the hypothesis, i.e., the predicted class of a sample \mathbf{x}_0 is chosen as the class which has the highest probability $P(y = k \mid \mathbf{x} = \mathbf{x}_0)$:

$$h(\mathbf{x}_0) = \operatorname{argmax}_k P(y = k \mid \mathbf{x} = \mathbf{x}_0) \tag{1.3}$$

Statistically speaking, this is the best possible classifier. The problem arises in the implementation. The exact probability density functions are unknown and have to be estimated from the available data, which is the source of error. Estimated version of the stated probabilities will be marked with a different symbols to stress out the fact they are just an estimates:

$$p_k(\mathbf{x}) := P(y = k \mid \mathbf{x}) \tag{1.4}$$

$$g_k(\mathbf{x}) := P(\mathbf{x} \mid y = k) \tag{1.5}$$

$$\pi_k := P(y = k) \tag{1.6}$$

Linear Discriminant Analysis estimates marginal probability term (π_k) as a ratio of number of samples belonging to class k and the total number of samples, whereas the class-conditional probability term in the Bayesian equation is estimated as a multivariate Gaussian function:

$$g_k(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} |\Sigma_k|^{1/2}} e^{-1/2(\mathbf{x}-\mu_k)^T \Sigma_k^{-1} (\mathbf{x}-\mu_k)} \tag{1.7}$$

, where m is the dimensionality of the feature space, i.e., number of features representing each sample. Function g_k is estimated class-conditional probability of class k , and μ_k and Σ_k are the mean and co-variance matrix for class k , respectively, and they are estimated from the available data as:

$$\mu_k = \frac{1}{n_k} \sum_i \mathbf{x}_i \Big|_{\forall \mathbf{x} \in k} \quad (1.8)$$

$$\Sigma_k = \frac{1}{n_k - K} \sum_i \left(\mathbf{x}_i - \mu_k \right) \left(\mathbf{x}_i - \mu_k \right)^T \Big|_{\forall \mathbf{x} \in k} \quad (1.9)$$

, where n_k represents the number of samples belonging to a class k . To simplify the model, LDA assumes that the co-variance matrices Σ_k are the same for all classes:

$$\Sigma_0 = \Sigma_1 = \dots = \Sigma_{K-1} = \Sigma \quad (1.10)$$

and they are usually calculated using the weighted average:

$$\Sigma = \frac{\sum_{k=1}^K n_k \Sigma_k}{\sum_{k=1}^K n_k} \quad (1.11)$$

The consequence of this assumption is the linearity of the decision boundary. Without this assumption the same calculus would lead to quadratic discriminant analysis, which has non-linear boundary.

In a two class example ($y \in \{0, 1\}$), all samples on the decision boundary will have the same probability of belonging to class 0 or 1:

$$D.B. = \left\{ \mathbf{x} \mid P(y = 0 \mid \mathbf{x} = \mathbf{x}_0) = P(y = 1 \mid \mathbf{x} = \mathbf{x}_0) \right\} \quad (1.12)$$

Following this idea, the decision boundary can be estimated by solving the equation:

$$\frac{g_0(\mathbf{x}) \pi_0}{\sum_{k=1}^K g_k \pi_k} = \frac{g_1(\mathbf{x}) \pi_1}{\sum_{k=1}^K g_k \pi_k} \quad (1.13)$$

$$\frac{1}{(2\pi)^{m/2} |\Sigma_0|^{1/2}} e^{-1/2(\mathbf{x}-\mu_0)^T \Sigma_0^{-1} (\mathbf{x}-\mu_0)} \pi_0 = \frac{1}{(2\pi)^{m/2} |\Sigma_1|^{1/2}} e^{-1/2(\mathbf{x}-\mu_1)^T \Sigma_1^{-1} (\mathbf{x}-\mu_1)} \pi_1 \quad (1.14)$$

If making the assumption on the equal co-variance matrices for both classes ($\Sigma_0 = \Sigma_1 = \Sigma$), and taking the logarithm, the equation takes the form:

$$-\frac{1}{2}(\mathbf{x} - \mu_0)^T \Sigma^{-1} (\mathbf{x} - \mu_0) + \log(\pi_0) = -\frac{1}{2}(\mathbf{x} - \mu_1)^T \Sigma^{-1} (\mathbf{x} - \mu_1) + \log(\pi_1) \quad (1.15)$$

, which can be written in the form of the linear function $\mathbf{x}^T \beta + \alpha = 0$ as:

$$\mathbf{x}^T (\Sigma^{-1} \mu_0 - \Sigma^{-1} \mu_1) + \frac{1}{2} (\mu_1^T \Sigma^{-1} \mu_1 - \mu_0^T \Sigma^{-1} \mu_0) + \log\left(\frac{\pi_0}{\pi_1}\right) = 0 \quad (1.16)$$

This equation represents the decision boundary between two classes, i.e., all samples lying on this line will have equal probability of belonging to class 0 and class 1. It should be noted that the slope of the line depends only on the class means and co-variance matrix, whereas a priori probabilities (which are the result of number of samples belonging to class 0 or 1) have effect only on the y -intercept term, i.e., the offset of the function. This is an interesting point that demands caution. If groups are unbalanced, that is, number of samples of one group is higher than in the other group, y -intercept of the decision boundary will be affected and the classifier will be biased by this disproportion. If groups are unbalanced because of the incomplete or missing data, whereas in reality they are balanced, this can have a negative effect.

When considering multiclass classification problem, probability of a sample belonging to each

class is firstly estimated by the equation:

$$p_k = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{x} - \mu_k)^T \Sigma^{-1} (\mathbf{x} - \mu_k) + \log (\pi_k) \quad (1.17)$$

and then the class is estimated as the one with the highest probability as:

$$h(\mathbf{x}) = \operatorname{argmax}_k p_k(\mathbf{x}) \quad (1.18)$$

Support Vector Machine

Try to solve the problem directly and never solve a more general problem as an intermediate step.

Vladimir Vapnik

Support vector machine is nowadays known as a very powerful classifier with a lot of different applications. The big advantage over LDA is the fact that it is a *non-parametric* classifier. The model is not obtained using assumptions of the form of the class density function and estimation of it's parameters, which is inevitably erroneous. Instead, SVM forms the decision boundary using the samples (not their density estimates) by maximizing the distance between samples and the boundary. This was the idea Vladimir Vapnik, the inventor of this method stood for. It is better to try to solve the problem directly and simply, without many intermediate steps that can be complicated and inaccurate.

In pattern recognition, the decision rule (h) is usually obtained by multiplying the sample (\mathbf{x}) by predefined weights (Θ):

$$\Theta^T \mathbf{x} + \Theta_0 \quad (1.19)$$

, where Θ_0 is a constant. If samples \mathbf{x}_0 and \mathbf{x}_1 lay on the decision boundary, following statements

are true:

$$\Theta^T \mathbf{x}_0 + \Theta_0 = \Theta^T \mathbf{x}_1 + \Theta_0 \quad (1.20)$$

$$\Theta^T (\mathbf{x}_0 - \mathbf{x}_1) = 0 \quad (1.21)$$

This result implies that Θ is perpendicular to the boundary:

$$\Theta \perp (\mathbf{x}_0 - \mathbf{x}_1) \quad (1.22)$$

The goal of the SVM is to find the decision boundary between two classes so that the distance between the samples and the decision boundary, i.e., the margin is maximized. The distance (d) from a sample to the decision boundary can be defined as the distance between the sample \mathbf{x} and any point lying on the boundary, \mathbf{x}_0 , projected onto the vector Θ .

$$d = \frac{\Theta^T (\mathbf{x} - \mathbf{x}_0)}{|\Theta|} \quad (1.23)$$

Term $|\Theta|$ is introduced to normalize the vector Θ . Without the normalization the distance would depend on the norm of Θ .

Since \mathbf{x}_0 is on the decision boundary, the expression $\Theta^T \mathbf{x}_0 + \Theta_0 = 0$ is valid, and, therefore, the expression for the distance can be written as:

$$d = \frac{\Theta^T \mathbf{x} + \Theta_0}{|\Theta|} \quad (1.24)$$

Margin (M) can be defined as the distance from the boundary to the closest sample:

$$M = \min_i d_i \quad (1.25)$$

Depending on which side of the boundary the sample is located, the distance can be positive or

negative. In order to keep it strictly positive, term y is introduced, where $y \in \{-1, 1\}$:

$$M = \min_i \{y_i d_i\} \quad (1.26)$$

$$M = \min_i \left\{ \frac{y_i (\Theta^T \mathbf{x}_i + \Theta_0)}{|\Theta|} \right\} \quad (1.27)$$

The objective is to maximize the margin M . Since Θ can be rescaled, a certain Θ exists so that $y_i (\Theta^T \mathbf{x}_i + \Theta_0) = 1$, which implies

$$\exists \Theta, y_i (\Theta^T \mathbf{x}_i + \Theta_0) = 1 \Rightarrow M = \min_i \left\{ \frac{1}{|\Theta|} \right\} \quad (1.28)$$

Therefore, to maximize the margin, a separating hyperplane should be found such that a norm of vector orthogonal to the hyperplane (Θ) is minimal.

For every point not on the boundary the following term is valid:

$$y_i (\Theta^T \mathbf{x}_i + \Theta_0) > 0 \quad (1.29)$$

Value C can be selected such that:

$$y_i (\Theta^T \mathbf{x}_i + \Theta_0) > C \quad (1.30)$$

$$y_i \left(\frac{\Theta^T \mathbf{x}_i}{C} + \frac{\Theta_0}{C} \right) > 1 \quad (1.31)$$

Since Θ and Θ_0 can be rescaled, it can be written:

$$\Theta := \frac{\Theta}{C}, \quad \Theta_0 := \frac{\Theta_0}{C} \quad (1.32)$$

$$y_i (\Theta^T \mathbf{x}_i + \Theta_0) > 1 \quad (1.33)$$

Finally the optimization problem states:

$$\min \frac{1}{2}|\Theta|^2, \quad s.t. \quad y_i (\Theta^T \mathbf{x}_i + \Theta_0) > 1. \quad (1.34)$$

L_2 norm is preferred because it has continuous derivative, whereas constant $1/2$ is introduced for the mathematical convenience. The optimization is solved using Lagrangian method as:

$$L(\Theta, \Theta_0, \alpha_i) = \frac{1}{2}|\Theta|^2 - \sum_{i=1}^n \alpha_i [y_i (\Theta^T \mathbf{x}_i + \Theta_0) - 1] \quad (1.35)$$

$$\frac{\partial L}{\partial \Theta} = \Theta - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \Theta = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (1.36)$$

$$\frac{\partial L}{\partial \Theta} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (1.37)$$

By rewriting the problem in 1.35 in terms of dual variable α , the following expression can be obtained:

$$L(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_j \sum_i \alpha_j \alpha_i y_j y_i \mathbf{x}_i^T \mathbf{x}_j \quad (1.38)$$

Since this function depends only on dual variable α , the solution can be obtained by maximization:

$$\max L(\alpha) \quad s.t. \quad \begin{cases} \alpha_i \geq 0 \\ \sum_i \alpha_i y_i = 0 \end{cases} \quad (1.39)$$

In this optimization problem, the objective has the form of quadratic function, whereas constraints are linear. This problem is typically solved using quadratic programming. Since it is a convex problem, the solution will always be global maximum. Once the dual variable α is found, the primal variable Θ can be calculated using the equation 1.36.

In the optimization, Karush-Kuhn-Tucker conditions need to be satisfied (Boyd and Vanden-

berghe, 2004). One of this condition is *complementary slackness*, stating that in the optimal point (the solution of the problem), the product of dual variable and the constraint must be zero:

$$\alpha_i [y_i (\Theta^T \mathbf{x}_i + \Theta_0) - 1] = 0 \quad (1.40)$$

This condition explains well the principle of SVM. Since the dual variable must be greater or equal to zero ($\alpha \geq 0$), there are two possibilities:

1. If α is greater than zero, $[y_i (\Theta^T \mathbf{x}_i + \Theta_0) - 1]$ must equal one:

$$\alpha_i > 0 \Rightarrow y_i (\Theta^T \mathbf{x}_i + \Theta_0) = 1 \quad (1.41)$$

2. If $[y_i (\Theta^T \mathbf{x}_i + \Theta_0) - 1]$ is greater than zero, α must be zero:

$$y_i (\Theta^T \mathbf{x}_i + \Theta_0) > 1 \Rightarrow \alpha = 0 \quad (1.42)$$

Since for all samples lying on the margin, the statement

$$y_i (\Theta^T \mathbf{x}_i + \Theta_0) = 1 \quad (1.43)$$

holds, α will be greater than zero only for the samples lying on the decision hyperplane, whereas for the samples further away from the hyperplane, α will be zero. Given the fact that Θ depends on linear combination of samples weighted by α (eq. 1.36), only the samples lying on the boundary will have effect in the calculation of Θ (where $\alpha > 0$), and they are called *support vectors*. This can be especially useful for the outlier samples.

The optimization problem does not depend on \mathbf{x} , but on $\mathbf{x}^T \mathbf{x}$. This allows the use of *kernel trick* and implicitly enables nonlinear transform of the feature space at little additional cost. Usually, non-linear decision boundary can be achieved by nonlinear transform of features:

$$\mathbf{x} \rightarrow \Phi(\mathbf{x}) \quad (1.44)$$

However, this operation is computationally expensive. The solution can be achieved using kernel

functions. Kernel is a function $K(x, y)$ for which:

$$K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y}) \quad (1.45)$$

Since in the equation 1.37 \mathbf{x} does not appear by itself, but in a form of dot product $\mathbf{x}^T \mathbf{x}$, non-linear transform can be used in a form of kernel trick:

$$L(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_j \sum_i \alpha_j \alpha_i y_j y_i K(\mathbf{x}_i, \mathbf{x}_j) \quad (1.46)$$

Most often used kernel is a radial basis kernel ($K_{RBF}(\mathbf{x}_i, \mathbf{x}_j)$):

$$K_{RBF}(\mathbf{x}_i, \mathbf{x}_j) = e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \quad (1.47)$$

Although SVM is conceptually designed as a two-class classifier, techniques for multiclass classification also exist, e.g. *one-versus-one* or *one-versus-all*.

Notes

■ prepisi od Monice	2
-------------------------------	---

Bibliography

- Boyd, S. and Vandenberghe, L. *Convex optimization*. Cambridge University Press, 2004.
- Clancy, E., Morin, E. L., and Merletti, R. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. *Journal of Electromyography and Kinesiology*, 12(1):1–16, 2002.
- Duchateau, J. and Enoka, R. M. Human motor unit recordings: Origins and insight into the integrated motor system. *Brain Research*, 1409:42–61, 2011.
- Englehart, K. and Hudgins, B. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 50(7):848–854, 2003.
- Farina, D., Jiang, N., Rehbaum, H., Holobar, A., Graimann, B., Dietl, H., and Aszmann, O. C. The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):797–809, 2014.
- Fougner, A., Scheme, E., Chan, A. D. C., Englehart, K., and Stavdahl, Ø. Resolving the Limb Position Effect in Myoelectric Pattern Recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(6):644–651, 2011.
- Fougner, A., Stavdahl, O., Kyberd, P. J., Losier, Y. G., and Parker, P. a. Control of upper limb prostheses: Terminology and proportional myoelectric control - A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(5):663–677, 2012.
- Freund, H. J., Büdingen, H. J., and Dietz, V. Activity of Single Motor Units from Human Forearm Muscles during Voluntary Isometric Contractions. *Journal of Neurophysiology*, 38(4):933–46, jul 1975.
- Hargrove, L., Englehart, K., and Hudgins, B. A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control. *Biomedical Signal Processing and Control*, 3(2):175–180, 2008.
- Hargrove, L. J., Englehart, K., and Hudgins, B. A comparison of surface and intramuscular myoelectric signal classification. *IEEE Transactions on Biomedical Engineering*, 54(5):847–853, 2007.
- Henneberg, K. Principles of Electromyography. In Bronzino, J., editor, *The Biomedical Engineering Handbook*. CRC Press, Boca Raton, second edition, 1999.
- Henneman, E., Somjen, G., and Carpenter, D. O. Functional Significance of Cell Size in Spinal Motoneurons. *Journal of Neurophysiology*, 28(3), 1965.

- Jiang, N., Dosen, S., Muller, K.-R., and Farina, D. Myoelectric Control of Artificial Limbs—Is There a Need to Change Focus? [In the Spotlight]. *IEEE Signal Processing Magazine*, 29(5): 152–150, 2012.
- Liddell, E. G. T. and Sherrington, C. S. Recruitment and some other Features of Reflex Inhibition. *Proceedings of the Royal Society of London B: Biological Sciences*, 97(686):488–518, 1925.
- Marateb, H. R., McGill, K. C., and Webster, J. G. Electromyographic (Emg) Decomposition. In *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley & Sons, Inc, 1999.
- Merletti, R. and Farina, D. *Surface Electromyography: Physiology, Engineering, and Applications*. Wiley-IEEE Press, Hoboken, New Jersey (USA), 2016.
- Merletti, R. and Parker, P. *Electromyography : physiology, engineering, and noninvasive applications*. Wiley-IEEE Press, 2004.
- Merletti, R., Avenaggiato, M., Botter, A., Holobar, A., Marateb, H., and Vieira, T. M. M. Advances in surface EMG: recent progress in detection and processing techniques. *Critical Reviews in Biomedical Engineering*, 38(4):305–45, 2010.
- Sherrington, C. S. Remarks on some Aspects of Reflex Inhibition. *Proceedings of the Royal Society of London B: Biological Sciences*, 97(686):519–545, 1925.
- Simon, A. M., Hargrove, L. J., Lock, B. a., and Kuiken, T. a. A decision-based velocity ramp for minimizing the effect of misclassifications during real-time pattern recognition control. *IEEE Transactions on Biomedical Engineering*, 58(8):2360–2368, 2011.
- Squire, J. *Muscle: design, diversity, and disease*. Benjamin/Cummnigs, Menlo Park, CA, USA, 1986.
- Staudenmann, D., Roeleveld, K., Stegeman, D. F., and van Dieen, J. H. Methodological aspects of SEMG recordings for force estimation - A tutorial and review. *Journal of Electromyography and Kinesiology*, 20(3):375–387, 2010.
- Tkach, D., Huang, H., and Kuiken, T. a. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of Neuroengineering and Rehabilitation*, 7:21, 2010.
- Young, A. J., Hargrove, L. J., and Kuiken, T. A. The Effects of Electrode Size and Orientation on the Sensitivity of Myoelectric Pattern Recognition Systems to Electrode Shift. *IEEE Transactions on Biomedical Engineering*, 58(9):2537–2544, 2011.
- Young, A. J., Smith, L. H., Rouse, E. J., and Hargrove, L. J. Classification of simultaneous movements using surface EMG pattern recognition. *IEEE Transactions on Biomedical Engineering*, 60(5):1250–1258, 2013.