Proc. 8th Intl. Conf. IEEE EMBC Neural Engineering Conference (NER'17), Shanghai, China (2017).

Use of Regularized Discriminant Analysis Improves Myoelectric Hand Movement Classification

Agamemnon Krasoulis, *Student Member, IEEE*, Kianoush Nazarpour, *Senior Member, IEEE* and Sethu Vijayakumar

Abstract— **Linear discriminant analysis (LDA) is the most commonly used classification method for movement intention decoding from myoelectric signals. In this work, we review the performance of various discriminant analysis variants on the task of hand motion classification. We demonstrate that optimal classification performance is achieved with regularized discriminant analysis (RDA), a method which generalizes various class-conditional Gaussian classifiers, including LDA, quadratic discriminant analysis (QDA), and Gaussian naive Bayes (GNB). The RDA method offers a continuum between these models via tuning two hyper-parameters which control the amount of regularization applied to the estimated covariance matrices. In this study, we performed a systematic classification performance comparison on four datasets. Hand motion was decoded from myoelectric and inertial data recorded from 60 able-bodied and 12 amputee subjects whilst they performed a range of 40 movements. We found that when the regularization parameters of the RDA classifier were carefully tuned via cross-validation, classification accuracy was statistically higher by a large margin as compared to any other discriminant analysis method (average improvement of 13.7% over LDA). Importantly, our findings were consistent across the able-bodied and amputee populations. This observation provides supporting evidence that our proposed methodology could improve the performance of pattern recognition-based myoelectric prostheses.**

I. INTRODUCTION

Pattern recognition-based myoelectric control aims at deciphering movement intention from biosignals to achieve naturalistic control of external devices, such as prosthetic hands and exoskeletons. In the myoelectric control literature, linear discriminant analysis (LDA) is the most extensively used method for motion classification for a number of reasons, including ease of implementation, computational efficiency, and performance robustness [1]–[3]. In this work, we review the efficacy of a general family of discriminant analysis and class-conditional Gaussian models in the context of hand motion classification.

*A. Krasoulis is supported in part by grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ac.uk/dtc) from the UK Engineering and Physical Sciences Research Council (EPSRC), UK Biotechnology and Biological Sciences Research Council (BBSRC), and the UK Medical Research Council (MRC). K. Nazarpour is supported by the EPSRC grants EP/M025594/1 and EP/M025977/1. S. Vijayakumar is supported by the Microsoft Research RAEng. Fellowship and EPSRC grant EP/H1012338/1.

A. Krasoulis is with the Institute for Adaptive and Neural Computation and the Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, UK A.Krasoulis@sms.ed.ac.uk

K. Nazarpour is with the School of Electrical and Electronic Engineering and the Institute of Neuroscience, Newcastle University, Newcastle, UK Kianoush.Nazarpour@newcastle.ac.uk

S. Vijayakumar is with the Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, UK sethu.vijayakumar@ed.ac.uk

II. BACKGROUND

Discriminant analysis is a family of supervised probabilistic models which assumes class-conditional multivariate Gaussian densities [4]. LDA is a special case of this family where a common covariance matrix, often referred to as the pooled covariance or within-class scatter matrix, is shared among classes. This assumption leads to linear decision boundaries (i.e. hyperplanes). In LDA, a test data point $\boldsymbol{x}_\star$ is assigned to the class c for which the linear discriminant function $\delta_c(\boldsymbol{x}_\star)$ is maximized:

$$\delta_c(\boldsymbol{x}_\star) = \boldsymbol{x}_\star^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_c - \frac{1}{2} \boldsymbol{\mu}_c^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_c + \log \pi_c, \qquad (1)$$

where π_c and $\boldsymbol{\mu}_c$, for $c = 1, \ldots, C$, are the prior probabilities and class means, respectively, $\boldsymbol{\Sigma}$ is the pooled covariance matrix, and C is the number of classes. The prior probabilities, mean vectors, and pooled covariance matrix can be estimated from the training data:

$$\hat{\pi}_c = \frac{N_c}{N} \qquad (2)$$

$$\hat{\boldsymbol{\mu}}_c = \frac{1}{N_c} \sum_{n \in c} \boldsymbol{x}_n \qquad (3)$$

$$\hat{\boldsymbol{\Sigma}} = \sum_{c=1}^{C} \sum_{n \in c} \frac{1}{N-C} (\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_c)(\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_c)^\top, \qquad (4)$$

where N_c is the number of training instances in class c, and N is the total number of training samples. The posterior probability for class c is given by the softmax function:

$$p(y = c | \boldsymbol{x}_\star) = \frac{e^{\delta_c(\boldsymbol{x}_\star)}}{\sum_{c'=1}^{C} e^{\delta_{c'}(\boldsymbol{x}_\star)}}. \qquad (5)$$

Quadratic discriminant analysis (QDA) is another type of class-conditional Gaussian model which does not make the LDA assumption (i.e. shared covariance matrix), thereby a separate covariance matrix has to be to estimated for each class. In this case, the decision boundaries are quadratic in feature space and the discriminant functions are given by:

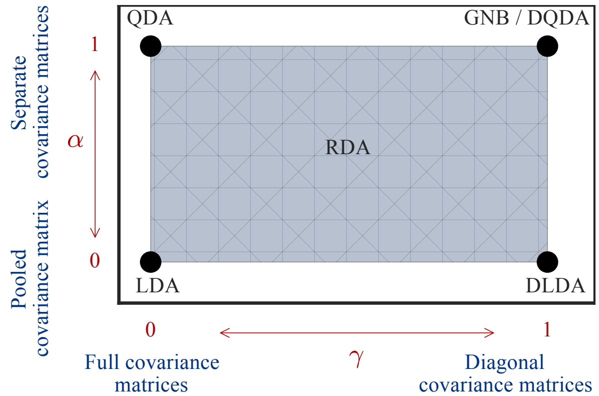

Fig. 1. Family of discriminant analysis classifiers. Classifiers such as LDA, QDA, DLDA and GNB / DQDA can be recovered as special cases of RDA via appropriate selection of regularization hyper-parameters α and γ.

$$\begin{aligned}\delta_c\left(\boldsymbol{x}_\star\right) = &-\frac{1}{2}\boldsymbol{x}_\star^\top \boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_\star + \boldsymbol{\mu}_c^\top \boldsymbol{\Sigma}_c^{-1}\boldsymbol{x}_\star \\ &-\frac{1}{2}\log|\boldsymbol{\Sigma}_c| + \log\pi_c - \frac{1}{2}\boldsymbol{\mu}_c^\top\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c,\end{aligned} \tag{6}$$

where $\boldsymbol{\Sigma}_c$ is the covariance matrix of class c, which can be estimated from the training data:

$$\hat{\boldsymbol{\Sigma}}_c = \frac{1}{N-1}\sum_{n\in c}\left(\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_c\right)^2 \tag{7}$$

Regularized discriminant analysis (RDA) is a method that generalizes LDA and QDA and provides a continuum of models between the two [5]. Similar to the QDA classifier, the class covariance matrices for this model are separate, however they are regularized toward the pooled covariance matrix and take the form:

$$\hat{\boldsymbol{\Sigma}}_c\left(\alpha\right) = \alpha\hat{\boldsymbol{\Sigma}}_c + \left(1-\alpha\right)\hat{\boldsymbol{\Sigma}}, \qquad 0\leq\alpha\leq 1. \tag{8}$$

The parameter α controls the amount of regularization. For $\alpha = 0$, we recover LDA, and for $\alpha = 1$, we recover QDA. A different form of regularization occurs when the estimated covariance matrices are regularized toward diagonal matrices, that is,

$$\hat{\boldsymbol{\Sigma}}\left(\gamma\right) = \left(1-\gamma\right)\hat{\boldsymbol{\Sigma}} + \gamma\,\mathrm{diag}(\hat{\boldsymbol{\Sigma}}) \tag{9}$$

and

$$\hat{\boldsymbol{\Sigma}}_c\left(\gamma\right) = \left(1-\gamma\right)\hat{\boldsymbol{\Sigma}}_c + \gamma\,\mathrm{diag}(\hat{\boldsymbol{\Sigma}}_c), \quad 0\leq\gamma\leq 1. \tag{10}$$

The two regularization approaches are orthogonal, and can be thus combined into:

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_c\left(\alpha,\gamma\right) = &\,\alpha\left(1-\gamma\right)\hat{\boldsymbol{\Sigma}}_c + \left(1-\alpha\right)\left(1-\gamma\right)\hat{\boldsymbol{\Sigma}} \\ &+\Big(\alpha\gamma\,\mathrm{diag}(\hat{\boldsymbol{\Sigma}}_c) + \left(1-\alpha\right)\mathrm{diag}(\hat{\boldsymbol{\Sigma}})\Big).\end{aligned} \tag{11}$$

The model described by Eq. (11) leads to a general family of models which treats as special cases various well-known classifiers including LDA, QDA, Gaussian naive Bayes (GNB), and diagonal linear discriminant analysis (DLDA), that is, an LDA model with diagonal pooled covariance matrix. A schematic representation of this family of models is shown in Fig. 1. It is worth noting, that the GNB model is occasionally referred to as diagonal quadratic discriminant analysis (DQDA).

Fig. 2. Sensor placement for an able-bodied and an amputee subject.

TABLE I
EXPERIMENTAL DATASETS

Dataset	Number of subjects	Medical condition	Sensing modalities
1	40	Able-bodied	sEMG, ACC
2	10	Transradial amputees	sEMG, ACC
3	20	Able-bodied	sEMG, ACC, GYRO, MAG
4	2	Transradial amputees	sEMG, ACC, GYRO, MAG

III. METHODS

A. Datasets

In this study, we used four datasets to evaluate and compare the performance of the family of classifiers introduced in the previous section. The first two datasets are publicly available, as part of the Ninapro project[1] (databases 2 and 3 in Atzori et al. [3]). The latter two datasets were collected by the authors by using the Ninapro protocol and made available on the same repository. One difference between the two pairs of datasets is that, for the first pair, the standard Delsys® Trigno™ sensors[2] were used which incorporate surface electromyogram (sEMG) electrodes and accelerometers. For the latter two datasets, the Delsys® Trigno™ IM system was used which also includes gyroscopes and magnetometers. The inertial measurement data (acceleration, angular velocity, magnetic field) were used in their raw format along with sEMG measurements to decode hand motion intention [6]. General information about all four datasets, including number of participants, medical conditions and input sensing modalities are presented in Table I.

In all experiments, eight sensors were equally placed around the subjects' forearm, and four sensors targeted the extensor digitorum communis, flexor digitorum superficialis, biceps, and triceps brachii muscles (Fig. 2). The participants sat in a chair and performed six repetitions of 40 movements which were instructed to them on a computer screen. The exercises included individuated finger, wrist, grasp and functional movements [3]. Myoelectric and inertial data were

[1] http://ninapro.hevs.ch
[2] http://www.delsys.com/

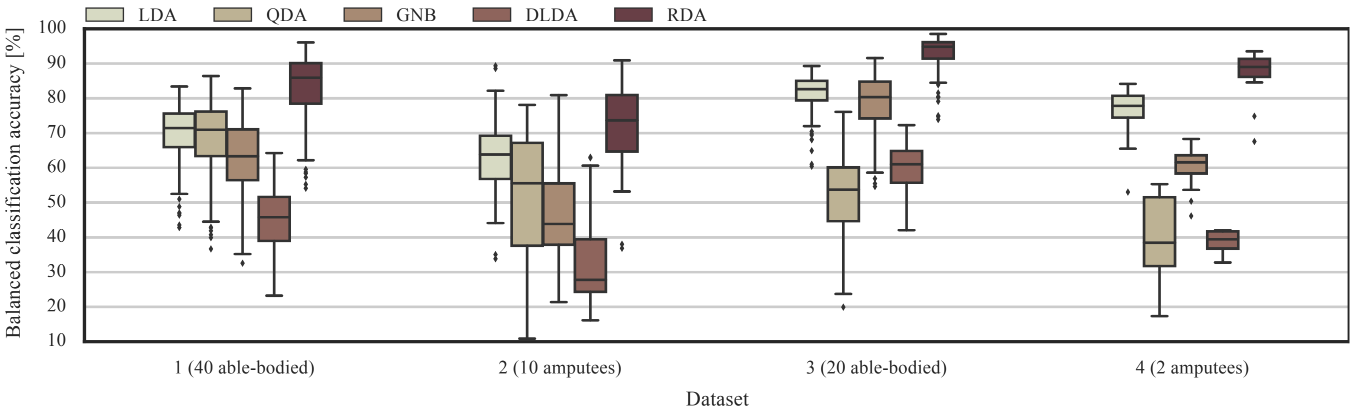

Fig. 3. Decoding performance comparison. The balanced classification accuracy is shown for five discriminant analysis variants and four datasets.

recorded at sampling frequencies of 2 KHz and 128 Hz, respectively.

B. Signal pre-processing and feature extraction

Myoelectric signals were digitally band-pass filtered in the range 20 Hz to 500 Hz. By using a shifting window approach, four time-domain sEMG features were extracted from each channel, namely the mean absolute value (MAV), waveform length (WL), 4^{th}-order auto-regressive (AR) and log-variance (Log-Var) [7]. The length of the window was set to 256 ms and the increment to 50 ms. For accelerometry (ACC) and inertial measurements (IMs), the mean value (MV) within the processing window was calculated. Thus, a total number of 10 features was extracted from each sensor for datasets 1 and 2 (7 sEMG, 3 ACC features), whereas the same figure for datasets 3 and 4 was 16 (7 sEMG, 9 IM features). Hence, the input feature space was 120- and 192-dimensional for the two pairs of datasets, respectively.

C. Cross-validation (CV) and hyper-parameter tuning

For each subject and movement, data from five repetitions were used to train models and the left-out repetition data were used to assess model performance. This procedure was repeated six times, hence resulting in 6-fold cross-validation (CV). To tune the regularization hyper-parameters α and γ for RDA, a grid search was performed in the range $[0, 1]$ with a step size of 0.05. In this case, inner-fold CV was used and the parameter values which yielded the highest average performance were selected.

D. Performance assessment

The class distribution of the test folds was balanced by removing a large proportion of the instances corresponding to the rest class. Decoding performance was then evaluated by using the standard classification accuracy (CA) metric.

IV. RESULTS

A performance comparison of five discriminant analysis classifiers (LDA, QDA, GNB, DLDA, RDA) is shown in Fig. 3. For all four datasets, RDA consistently outperformed all other classifiers and it was followed by LDA, GNB, QDA and

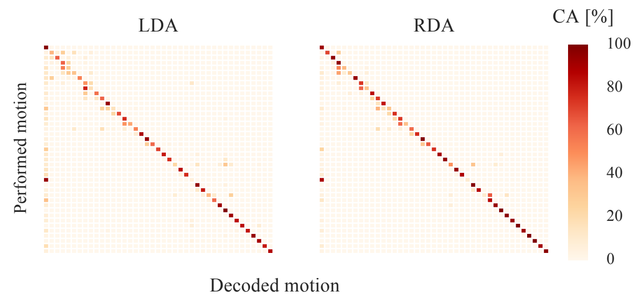

Fig. 4. LDA and RDA performance comparison. Representative confusion matrices are shown for an amputee subject (dataset 4). Color map represents balanced classification accuracy scores.

DLDA. The average CA median difference between RDA and LDA was 13.7%. All performance differences between classifiers were statistically significant ($p < 10^{-3}$, Friedman Test followed by pair-wise Wilcoxon rank-sum tests with Šidák correction for multiple comparisons). Representative confusion matrices for one amputee subject (dataset 4) are shown in Fig. 4 for LDA and RDA.

The distribution of hyper-parameters α and γ for RDA as selected by CV is shown in Fig. 5. The optimal selection for γ was almost consistently 0 (with very few exceptions where it was 0.05), whereas for α it was in the range 0.15 to 1.

V. DISCUSSION

A. Comparison of models, overfitting and regularization

The RDA classifier consistently outperformed all other models. This was expected, since the RDA model is flexible and can treat all other models as special cases (Eq. 11, Fig. 1). The two hyper-parameters of the RDA classifier were tuned such that the cross-validated CA was maximized, therefore, it was guaranteed that its performance would be at least as good as that of any other model.

The LDA model assumption, that is, classes share a common covariance matrix, is very strong and fundamentally wrong. One would expect that QDA should outperform LDA since it is more flexible and does not make this assumption. The reason why this is not often the case is that QDA is

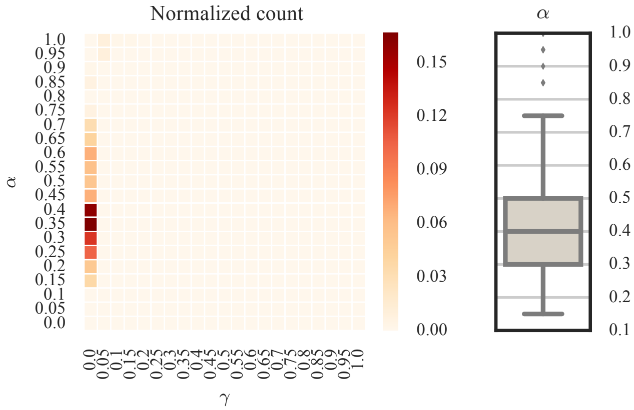

Fig. 5. Normalized count of selected values for RDA parameters α and γ. The optimal value for γ was almost consistently 0 (with very few exceptions where it was 0.05), whereas for α it varied in the range 0.15 to 1. The marginal distribution of α selected values is shown separately on the right.

heavily prone to overfitting. The number of free parameters that have to be estimated in the general class-conditional Gaussian model is $C\left(D+1\right)D/2$, where C is the number of classes and D is the dimensionality of the feature space. In our case, the number of free parameters was approximately 2.86×10^5 for datasets 1 and 2, and 7.6×10^5 for datasets 3 and 4. Taking into consideration that a typical CV fold included on average 3.6×10^3 training samples, it is obvious that this method suffered profoundly from overfitting. It is, therefore, not surprising that the classification performance of QDA for datasets 3 and 4 was inferior to that for datasets 1 and 2 (Fig. 3), since in the former case the feature space was larger (Section III-B). Conversely, the performance of the other classifiers was improved, as expected, when the additional sensory modalities (gyroscopes and magnetometers) were included in the set of features [6].

In the limit of infinite amount of data one should expect that QDA would always outperform LDA. In practice, however, it is not feasible to collect vast quantities of training data, especially with amputees. The benefit of using RDA lies in that it can make use of the theoretical advantage of QDA over LDA without being susceptible to overfitting, as a result of regularizing the class covariance matrices toward the pooled covariance matrix (α hyper-parameter).

The γ hyper-parameter is used in the RDA model to introduce a different form of regularization, that is, it shrinks the estimated covariance matrices toward diagonal matrices. In the extreme case (i.e. $\gamma = 1$). the naive Bayes model is recovered which assumes class-conditional feature independence. Nevertheless, such behaviour should neither be desired nor expected, since many features originate from the same measurements (i.e. we extract multiple features from the same sEMG signals). Input features which do not stem from the same measurements are still expected to exhibit strong correlations when they are recorded in nearby locations on the skin surface, due to signal crosstalk [8]. For that reason, it should not be surprising that the optimal value for γ was almost consistently 0 (Fig. 5).

B. *Computational and memory requirement considerations*

One strong advantage of the LDA model is that decision boundaries are linear in feature space. Consequently, the time complexity of assigning class probabilities to a test sample is $O\left(CD\right)$, that is, it scales linearly with the feature dimensionality. The space complexity for LDA is also $O\left(CD\right)$. For general class-conditional Gaussian models like QDA and RDA that are implemented efficiently, that is, when inverse covariance matrices are precomputed and stored in memory, the computational complexity is $O\left(CD^2\right)$ and space complexity is $O\left(CD^3\right)$. For small to medium-sized feature spaces, this would not pose a problem for real-time implementations.

VI. CONCLUSIONS

In this study, we reviewed and performed a systematic comparison of discriminant analysis models on hand motion classification. We demonstrated that a large improvement in CA can be achieved, both for able-bodied and amputee subjects, by using the RDA algorithm, as compared to the commonly used LDA model. The regularizing parameters of the RDA model can be easily tuned via hold-out or cross-validation.

Our study suggests that pattern recognition-based myoelectric control systems have the potential to benefit from deploying the RDA method in the decoding stage. Further verification of our results during real-time control paradigms is required and currently seen as a future research direction.

REFERENCES

[1] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, “Classification of simultaneous movements using surface EMG pattern recognition,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1250–1258, 2013.

[2] E. Scheme and K. Englehart, “Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use,” *The Journal of Rehabilitation Research and Development*, vol. 48, no. 6, p. 643, 2011.

[3] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” *Sci. Data*, vol. 1, p. 140053, 2014.

[4] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics, Springer, Berlin, 2001, vol. 1.

[5] J. H. Friedman, “Regularized discriminant analysis,” *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[6] I. Kyranou, A. Krasoulis, M. S. Erden, K. Nazarpour, and S. Vijayakumar, “Real-Time classification of multi-modal sensory data for prosthetic hand control,” in *6th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, June 2016, pp. 536–541.

[7] J. M. Hahne, F. Biebmann, N. Jiang, H. Rehbaum, D. Farina, F. C. Meinecke, K.-R. Muller, and L. C. Parra, “Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control.” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 22, no. 2, pp. 269–279, 2014.

[8] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O. Aszmann, “The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 22, no. 4, pp. 797–809, July 2014.