Using PCA on EEG Data to Distinguish Sleep Stages

Ida Hönigmann
Technical University Vienna, Austria
Email: e12002348@student.tuwien.ac.at

***Abstract*—[TODO]**

I. Introduction

[TODO general introduction]

A. EEG Data and Sleep Stages

Ganong [2] describes typical patterns observed in electroencephalogram (EEG) data of a sleeping person. He describes the EEG patterns associated with rapid eye movement (REM) sleep and non-REM (NREM) sleep.

NREM sleep is further partitioned into four (although some only use three) stages, termed Stage 1 (S1) to Stage 4 (S4). Example EEG data of these different sleep stages can be seen in Figure **??** [TODO image]. The EEG data of these stages is characterized as follows:

S1: low-amplitude, high-frequency
S2: appearance of sleep spindles (bursts of higher amplitude, lower frequency waves)
S3: increased amplitude, lower frequency
S4: maximal amplitude, minimal frequency

In REM sleep the EEG data is that of high frequency and low amplitude patterns, resembling the data observed in alert humans.

II. Study of Literature

A substantial body of scientific research has been devoted to exploring Principal Component Analysis (PCA). The foundation of this method was laid by Pearson [9] and Hotelling [4].

An introduction to PCA, as well as a good overview on how to derive the formula used to compute the Principal Components (PC) is given by Shlens [11]. Recent applications and variants of PCA are explored by Jolliffe et. al. [6].

Shlens discusses the limitations of PCA, as well as examples in which PCA fails [11], such as the requirement of linearly dependent data. Tenenbaum proposes a non-linear method to combat this problem[12].

Generally speaking the variables must not have third or higher order dependencies[1] between them. In some cases it is possible to reduce a problem with higher order dependencies to a second order one by applying a non-linear transformation beforehand. This method is called kernel PCA[11].

Another method for combating this problem is Independent Component Analysis (ICA) which is discussed by Naik et. al.[8].

The given problem of distinguishing sleep stages given some EEG data has been investigated by use of PCA, as well as neural networks. Some of these works are summarized below.

A review of different methods in the preprocessing, feature extraction and classification is given by Boostani et. al.[1]. They find that using a random forest classifier and entropy of wavelet coefficients as feature gives the best results.

Tăuţan et. al.[13] compare different methods of dimensionality reduction on EEG data, such as PCA, factor analysis and autoencoders. They conclude that PCA and factor analysis improves the accuracy of the model.

Putilov[10] used PCA to find boundaries between Stage 1, Stage 2 and Stage 3. Changes in the first two PC were related to changes between the Stage 1 and Stage 2, while changes in the fourth PC exhibited a change in sign at the boundary of Stage 2 and Stage 3. This suggests that changes between Stage 1 and Stage 2 are easier to detect that ones between Stage 2 and Stage 3.

Metzner et. al.[7] try to rediscover the different human-defined sleep stages. They find that using PCA on the results makes clusters apparent. These clusters could then be used as a basis for a redefinition of sleep stages.

The PhysioNet/Computing in Cardiology Challange 2018 was a competition using a similar data[3]. The goal was to identify arousal during sleep from EEG, EOG, EMG, ECG and SaO2 data given. The winning paper of this competition describes the use of a dense recurrent convolutional neural network (DRCNN) consisting of multiple dense convolutional layers, a bidirectional long-short term memory layer and a softmax output layer[5].

As shown in this section, the utilization of PCA to analyze EEG data has been used with success.

III. Mathematical Basics

We define mathematical notation, which will be used in Section IV to define the PCA.

A. Covariance

Assume we have two sets of n observations of variables with mean 0. Let us call the first list of observations $\mathbf{a} = (a_1, ..., a_n)$

[1] e.g. $\mathbb{E}[x_i x_j x_k] \neq 0$ for some i, j, k assuming mean-free variables

and the second $\mathbf{b} = (b_1, ..., b_n)$.

Definition 1 (covariance). *Let us define the covariance of* $\mathbf{a} \in \mathbb{R}^n$ *and* $\mathbf{b} \in \mathbb{R}^n$ *as*

$$\sigma_{\mathbf{ab}} := \frac{1}{n}\sum_{i=1}^{n} a_i b_i = \frac{1}{n}\mathbf{a}\cdot\mathbf{b}^T.$$

From the definition it is obvious that the covariance is symmetric, $\sigma_{\mathbf{ab}} = \sigma_{\mathbf{ba}}$. In the special case $\mathbf{a} = \mathbf{b}$ the covariance $\sigma_{\mathbf{aa}}$ is called *variance* $\sigma_{\mathbf{a}}^2$.

Definition 2 (covariance matrix). *Generalizing to* m *variables* $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_m}]$*, each having been observed* n *times, gives us the covariance matrix.*

$$\mathbf{C_X} := \begin{pmatrix} \sigma_{\mathbf{x_1x_1}} & \cdots & \sigma_{\mathbf{x_1x_m}} \\ \vdots & \ddots & \vdots \\ \sigma_{\mathbf{x_mx_1}} & \cdots & \sigma_{\mathbf{x_mx_m}} \end{pmatrix} = \frac{1}{n}\mathbf{XX}^T$$

The covariance matrix is a symmetric $m \times m$ matrix.

B. Diagonalizable Matrix

Definition 3 (Diagonalizable Matrix). *A square matrix* $\mathbf{A}$ *is called diagonalizable, if there exists a invertable matrix* $\mathbf{P}$ *and a diagonal matrix* $\mathbf{D}$ *such that* $\mathbf{A} = \mathbf{PDP}^{-1}$.

Definition 4 (Symmetric matrix). *A square matrix* $\mathbf{A}$ *is called symmetric, if* $\mathbf{A}^T = \mathbf{A}$.

Theorem 1. *Every symmetric matrix is diagonalizable.*

The proof of this theorem requires some preparation, which we will do now.

Definition 5 (Eigenvalues and Eigenvectors). *Let* $\mathbf{A}$ *be a real* $m \times m$ *matrix.* $\lambda \in \mathbb{R}$ *is called a eigenvalue with eigenvector* $\mathbf{v} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ *if*

$$\mathbf{Av} = \lambda\mathbf{v}. \tag{1}$$

Lemma 1. *Every square* $m\times m$ *matrix has* m *(not necessarily unique) eigenvalues.*

Proof. We can rewrite equation 1 as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$$

This allows us to interpret $(\mathbf{A} - \lambda\mathbf{I})$ as a function, which takes vectors $\mathbf{v} \in \mathbf{R}^m$. For λ to be a eigenvalue of $\mathbf{A}$ with eigenvector $\mathbf{v}$ it has to satisfy $\mathbf{v} \in \ker(\mathbf{A} - \lambda\mathbf{I})$ and $\mathbf{v} \neq 0$. From this we gather that all λ with $\ker(\mathbf{A} - \lambda\mathbf{I}) \neq \{\mathbf{0}\}$ are eigenvalues. We know that this holds if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. The determinant is a polynomial of degree m which can be expressed in the form $(\lambda - \lambda_1)...(\lambda - \lambda_m)$ with $\lambda_1, ..., \lambda_m \in \mathbb{C}$. These $\lambda_1, ..., \lambda_m$ are the m eigenvalues we wanted to find. $\square$

Lemma 2. *A symmetric matrix has real eigenvalues.*

Proof. Let $\bar{\cdot}$ denote the complex conjugate. Define a complex dot product

$$(\mathbf{u}, \mathbf{v}) := \sum_{i=1}^{m} u_i \bar{v}_i$$

This dot product has the following properties for all $\mathbf{A} \in \mathbb{C}^{m\times m}, \mathbf{u}, \mathbf{v} \in \mathbb{C}^m, \lambda \in \mathbb{C}$

- $(\mathbf{Au}, \mathbf{v}) = (\mathbf{u}, \mathbf{A}^T\mathbf{v})$,
- $(\lambda\mathbf{u}, \mathbf{v}) = \lambda(\mathbf{u}, \mathbf{v})$,
- $(\mathbf{u}, \lambda\mathbf{v}) = \bar{\lambda}(\mathbf{u}, \mathbf{v})$
- $(\mathbf{u}, \mathbf{u}) = 0 \iff \mathbf{u} = 0$

Let $\mathbf{A}$ be a symmetric matrix with eigenvalue $\lambda \in \mathbb{C}$. From this it follows that for all $\mathbf{u} \in \mathbb{C}^m$

$$\lambda(\mathbf{u}, \mathbf{u}) = (\lambda\mathbf{u}, \mathbf{u}) = (\mathbf{Au}, \mathbf{u}) = (\mathbf{u}, \mathbf{A}^T\mathbf{u}) = (\mathbf{u}, \mathbf{Au}) = (\mathbf{u}, \lambda\mathbf{u}) = \bar{\lambda}(\mathbf{u}, \mathbf{u}).$$

We derive that $\lambda = \bar{\lambda}$ and thus $\lambda \in \mathbb{R}$. $\square$

Lemma 3. *The eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal.*

Proof. Let λ_1, λ_2 be two distinct eigenvalues with eigenvectors $\mathbf{v}_1, \mathbf{v}_2$ of the matrix $\mathbf{A}$.

$$\lambda_1\mathbf{v}_1 \cdot \mathbf{v}_2 = (\lambda_1\mathbf{v}_1)^T\mathbf{v}_2 = (\mathbf{Av}_1)^T\mathbf{v}_2 = \mathbf{v}_1^T\mathbf{A}^T\mathbf{v}_2 = \mathbf{v}_1^T\mathbf{Av}_2 = \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) = \lambda_2\mathbf{v}_1 \cdot \mathbf{v}_2$$

This shows that $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and as λ_1 and λ_2 are distinct, $\mathbf{v}_1$ and $\mathbf{v}_2$ must be orthogonal. $\square$

Now we have everything we need to prove theorem 1.

Proof of Theorem 1. Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ be a symmetric matrix. From lemma 1 we know that eigenvalues $\lambda_1, ..., \lambda_m$ with corresponding eigenvectors $\mathbf{v}_1, ..., \mathbf{v}_m$ exist.

Define the following matrices

$$\mathbf{D} := \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} \quad \mathbf{E} := \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{pmatrix}$$

In order to show that $\mathbf{A} = \mathbf{EDE}^{-1}$ we calculate

$$\mathbf{AE} = \begin{pmatrix} \mathbf{Av}_1 & \cdots & \mathbf{Av}_m \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_m\mathbf{v}_m \end{pmatrix} = \mathbf{ED}.$$

From lemma 3 we know that the eigenvectors, and therefore the columns of $\mathbf{E}$, are orthogonal. From this it follows that $rank(\mathbf{E}) = m$ which gives us the existence of $\mathbf{E}^{-1}$.

This shows that $\mathbf{A}$ is diagonalizable. $\square$

Lemma 4. *If the columns of a matrix* $\mathbf{A}$ *are orthogonal, then* $\mathbf{A}^{-1} = \mathbf{A}^T$.

Proof. Let $(\mathbf{a}_i)_{i=1,...,m}$ be the columns of the matrix.

$$\forall i, j : \mathbf{a}_i^T\mathbf{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \implies \mathbf{A}^T\mathbf{A} = \mathbf{I}$$

This shows $\mathbf{A}^{-1} = \mathbf{A}^T$. $\square$

IV. Principal Component Analysis

V. Sleep Stages and EEG Data

VI. Data and Algorithm

1) subdivide eeg signals in the temporal domain
2) apply fft transforming into frequency domain
3) pca
4) achive dimensinality reduction
5) classification of sleep stages
6) visulisation

VII. Results

VIII. Conclusion

References

[1] Reza Boostani, Foroozan Karimzadeh, and Mohammad Nami. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine*, 140:77 – 91, 2017. Cited by: 212.

[2] William F. Ganong. *Review of medical physiology*. Appleton & Lange, Stamford, Conn, 18. ed edition, 1997.

[3] Mohammad M Ghassemi, Benjamin E Moody, Li wei H Lehman, Christopher Song, Qiao Li, Haoqi Sun, Roger G Mark, M Brandon Westover, and Gari D Clifford. You snooze, you win: the physionet/computing in cardiology challenge 2018. *2018 Computing in Cardiology Conference (CinC)*, pages 1–4, 2018.

[4] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[5] Matthew Howe-Patterson, Bahareh Pourbabaee, and Frederic Benard. Automated detection of sleep arousals from polysomnography data using a dense convolutional neural network. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.

[6] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Royal Society*, 374(2065), 2016.

[7] Claus Metzner, Achim Schilling, Maximilian Traxdorf, Holger Schulze, Konstantin Tziridis, and Patrick Krauss. Extracting continuous sleep depth from eeg data without machine learning. *Neurobiology of Sleep and Circadian Rhythms*, 14, 2023. All Open Access, Gold Open Access, Green Open Access.

[8] Ganesh R Naik and Dinesh K Kumar. An overview of independent component analysis and its applications. *Informatica*, 35(1), 2011.

[9] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[10] Arcady A. Putilov. Principal component analysis of the eeg spectrum can provide yes-or-no criteria for demarcation of boundaries between nrem sleep stages. *Sleep Science*, 8(1):16–23, 2015.

[11] Jonathon Shlens. A tutorial on principal component analysis. 2014.

[12] J.B. Tenenbaum, V. De Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319 – 2323, 2000. Cited by: 10812.

[13] Alexandra-Maria Tăuţan, Alessandro C. Rossi, Ruben de Francisco, and Bogdan Ionescu. Dimensionality reduction for eeg-based sleep stage detection: comparison of autoencoders, principal component analysis and factor analysis. *Biomedical Engineering / Biomedizinische Technik*, 66(2):125–136, 2021.