KTH Royal Institute of Technology

Department of Mathematics

Project Report

Facial Expression Recognition Using Topological Data Analysis

Yuqi Shao, Lea Keller, Maxime Scali, Niclas Popp

Course: SF2956 Topological Data Analysis

Teacher: Wojciech Chachólski, Andrea Guidolin

Submission Date: 15.11.2021

Abstract

Facial expression is one of the most direct ways for humans to communicate reactions and intentions. Yet, it remains difficult to automatically identify emotions in images of human faces. In this project, we propose an approach based on topological data analysis to address this question. For this purpose we make use of the pictures contained in the Facial Expression 2013 (FER2013) data set that features 7 emotion categories. We calculate relevant topological features and subsequently perform pairwise classification using support vector machines with a custom kernel. However, we only achieve moderate classification accuracy in the end.

Contents

1 Introduction **1**

2 Mathematical Foundations **2**
2.1 Topological Data Analysis . . . 2
2.1.1 Distances . . . 2
2.1.2 Dendrograms . . . 3
2.1.3 Parametrised Vector Space . . . 3
2.1.4 Simplicial Complexes . . . 3
2.1.5 Homology . . . 4
2.1.6 Piece-Wise Constant Functions . . . 4
2.1.7 General Pipeline . . . 4
2.1.8 Stability of the Process . . . 5
2.1.9 Hierarchical Clustering . . . 6
2.2 Support Vector Machines . . . 6
2.2.1 Primal Problem . . . 6
2.2.2 Kernel Trick . . . 7

3 Methods **8**
3.1 Dataset . . . 8
3.2 Topology on the Level of Emotions . . . 8
3.3 Topology on the Level of Images . . . 9

4 Results **11**
4.1 Embedding 1 . . . 11
4.2 Embedding 2 . . . 12
4.3 Classification Using Support Vector Machines . . . 14

5 Conclusion **15**

6 Supplementary Figures **16**

1 Introduction

Facial expression recognition (FER) plays an important role for the communication between humans. Facial expression is one of the most immediate ways to express intentions and emotions. Apart from that, FER has various promising applications in science and technology, for instance in traffic safety, virtual reality and cognitive science. Consequently, research on the detection and classification of facial expressions has emerged as a major field over the last two decades [3]. Mainly the advances in computer vision and deep learning have accelerated this process. Nevertheless, robust FER in uncontrolled environments still remains an open problem due to disturbances such as change in occlusion, illumination, and noise. In this project we aim to investigate if tools from topological data analysis (TDA) can provide advances to this question.

We selected the Facial Expression 2013 (FER2013) data set [2] as basis for our analysis. It features over $35,000$ grayscale images from 7 different emotions: *happiness, neutral, sadness, anger, surprise, disgust, fear*. These images were collected for a classification challenge on Kaggle. Human accuracy is estimated to be around 68% which illustrates the difficulty of the FER problem.
For the TDA based approach presented in this report we make use of two embeddings for the data. The first embedding allows us to analyse the topology on the level of emotions. Each emotion is represented by a point cloud where a data point corresponds to a picture that displays the respective expression. We analyse this structure using hierarchical clustering, bar codes and stable ranks based on H_0 and H_1 homologies to assess the similarities between the different emotions.
Subsequently, we make use of a second embedding that enables us to investigate the topological features on the level of images. Each picture is represented by a point cloud of the grayscale values of each pixel combined with its position. We again calculate the stable ranks based on the H_0 and H_1 homologies together with the mean and standard deviation per emotion.
Using this second embedding we aim to perform pairwise classification of images for each possible pair of two categories. Since the theory of stable ranks facilitates the application of kernel-based machine learning algorithms together with TDA, we make use of support vector machines (SVM) with a custom kernel to perform this task.

The report for this project is divided into 5 parts. For better readability all necessary mathematical foundations are introduced in chapter 2. The methods for applying this theory to the data set are discussed in chapter 3. In chapter 4 we present our main results. Finally, we conclude with a discussion about how our findings can be interpreted and put into context.

2 Mathematical Foundations

2.1 Topological Data Analysis

2.1.1 Distances

Definition 2.1. Let X be a set. A *distance* on X is a function $d : X \times X \longrightarrow [0, \infty]$ that satisfies

- Symmetry : $d(x, y) = d(y, x)$
- Reflexivity : $d(x, x) = 0$

for all $x, y \in X$. d is called a *pseudometric* if in addition the triangle inequality holds.

$$d(x, y) + d(y, z) \geq d(x, z), \forall x, y, z \in X. \tag{1}$$

If furthermore $d(x, y) = 0 \Longleftrightarrow x = y$, then d is called a *metric* on X. We denote the set of all distances on X by $\mathcal{D}(X)$.

Example 2.2. The main distances on $\mathbb{R}^n$ used in this project are the l_p *metric*

$$d_p(x, y) = (|x_1 - y_1|^p + ... + |x_n - y_n|^p)^{1/p} \tag{2}$$

and in particular the *Cityblock metric*

$$d_1(x, y) = |x_1 - y_1| + ... + |x_n - y_n| \tag{3}$$

Based on a distance space (X, d), we define the partition $(X, X/d_t)$ which is induced by the equivalence relation $\sim$, where $x \sim y \Longleftrightarrow d(x, y) \leq t$.

The symbol $\Delta[X]$ denotes the set of all non-empty finite subsets of X. We want to extend a distance d on X to a distance on $\Delta[X]$.

Definition 2.3. An *extension* along $X \subset \Delta[X]$ is a function $\psi : \mathcal{D}(X) \to \mathcal{D}(\Delta[X])$ such that $\psi d(\{x\}, \{y\}) = d(x, y)$.

Example 2.4. Let A, B be non-empty finite subsets of X and d a distance on X. Examples of extensions of d are single linkage (sd), complete linkage (cd) and average linkage (ad):

$$sd(A, B) = \min\{d(a, b) \mid a \in A,\ b \in B\}$$

$$cd(A, B) = \begin{cases} \max\{d(a, b) \mid a \in A,\ b \in B\} & \text{if } A \neq B \\ 0 \quad \text{if } A = B \end{cases}$$

$$ad(A, B) = \begin{cases} \frac{\sum_{a \in A, b \in B} d(a,b)}{|A|\,|B|} & \text{if } A \neq B \\ 0 \quad \text{if } A = B \end{cases}$$

2.1.2 Dendrograms

Definition 2.5. A *dendrogram* is sequence of sets D_t indexed by $t \in [0, \infty)$ together with a sequence of functions $D_{s<t} : D_s \to D_t$ indexed by pairs $s < t$. These functions are required to satisfy the following propreties:

- For all $0 \leq r < s < t < \infty$, the following diagram commutes:

$$\begin{array}{ccc} D_r & \xrightarrow{D_{r<s}} & D_s \\ & \searrow^{D_{r<t}} & \downarrow{\scriptstyle D_{s<t}} \\ & & D_t \end{array}$$

- There is a finite sequence of real numbers $0 = a_0 < ... < a_n$ such that if both s and t belong to the same interval among $[a_0, a_1), ..., [a_{n-1}, a_n), [a_n, \infty)$, then $D_{s<t}$ is a bijection.

- The functions $D_{s<t}$ are surjectives for all $s < t$.

2.1.3 Parametrised Vector Space

Definition 2.6. Let K be a field. A *tame vector space parametrised by* $[0, \infty)$ is a sequence of finite dimensional K-vector spaces V_t indexed by $t \in [0, \infty)$, together with a sequence of linear functions $V_{s<t} : V_s \to V_t$ indexed by pairs $s < t$ that satisfy:

- For all $0 \leq r < s < t < \infty$ the following diagram commutes:

$$\begin{array}{ccc} V_r & \xrightarrow{V_{r<s}} & V_s \\ & \searrow^{V_{r<t}} & \downarrow{\scriptstyle V_{s<t}} \\ & & V_t \end{array}$$

- There is a finite sequence of real numbers $0 = a_0 < ... < a_n$ such that if both s and t belong to the same interval among $[a_0, a_1), ..., [a_{n-1}, a_n), [a_n, \infty)$, then $V_{s<t}$ is an isomorphism.

Definition 2.7. Let $0 \leq a < b \leq \infty$. A *bar* $[a, b)_t$ is a tame vector space defined by

$$[a, b)_t = \begin{cases} K & \text{if } t \in [a, b) \\ 0 & \text{otherwise} \end{cases}$$

where $[a, b)_{s<t}$ is the identity function if $s, t \in [a, b)$ and 0 otherwise.

2.1.4 Simplicial Complexes

The purpose of TDA is to endow our space X with a structure that enables us to use topological tools on it. A useful approach for this purpose starts with simplicial complexes.

Definition 2.8. A collection K of some finite non empty subsets of X is called a *simplicial complex* if for $\sigma \in K$ it holds $\omega \subset \sigma \Longrightarrow \omega \in K$. An element of K of cardinality n is called a simplex of dimension $n-1$.

A special kind of simplicial complexes are the so called Vietoris-Rips complexes.

Definition 2.9. Let (X, d) be a distance space. The *Vietoris-Rips complex* of (X, d) at scale $t \geq 0$, denoted $VR_t(d)$ is given by the collection of subsets $\{x_1, ..., x_n\}$ such that $d(x_i, x_j) \leq t$ for all i and j in $\{1, ..., n\}$.

2.1.5 Homology

An useful tool to perform geometric and topological data analysis is homology.

Definition 2.10. The *homology* H_n or the *n-th homology* of a simplicial complex $K \subset \text{Part}(X)$ is the vector space defined by $H_n(X, K) = ker(\delta_n)/im(\delta_{n+1})$.

δ_n is the so called *boundary function*. For a more in depth elaboration of this concept we refer to [6].
δ_n maps a set of dimension n on its boundaries of dimensions $n-1$. H_n can therefore be seen as the vector space generated by the n-dimensional holes in K.

2.1.6 Piece-Wise Constant Functions

Definition 2.11. A function $f : [0, \infty] \longrightarrow (-\infty, \infty)$ is called *piece-wise constant* (PCF) if there is a finite number of elements $0 = a_0 < a_1 < ... < a_n$ in $[0, \infty)$ for which the restrictions of f to the right open intervals $[a_0, a_1), ..., [a_{n-1}, a_n), [a_n, \infty)$ are constant functions.

Our results are mainly represented through PCFs. In order to compare such functions, we introduce a distance on the set of PCFs:

Definition 2.12. The *interleaving distance* is given by

$$d_{\bowtie}(f, g) = \begin{cases} \infty \quad \text{if } \{v \mid f(x) \geq g(x+v) \text{ and } g(x) \geq f(x+v) \text{ for any } x\} = \emptyset \\ \inf\{v | f(x) \geq g(x+v) \text{ and } g(x) \geq f(x+v) \text{ for any } x\} \text{ otherwise} \end{cases} \tag{4}$$

2.1.7 General Pipeline

The main purpose of TDA is to transform data in representations that are easier to analyse than distance spaces. Given a pseudometric space (X, d) we obtain the corresponding Vietoris-Rips complex. This choice is justified by the fact that such complexes encode the entirety of the information contained in (X, d).

Proposition 2.13. *Let d and d' be distances on X. Then $d = d' \Longleftrightarrow VR_t(d) = VR_t(d')$ for all t in $[0, \infty)$.*

Since the pseudometric spaces (X, d) that we are interested in are finite, there is a finite number of steps $t_1, ..., t_n$ such that if t, $t' \in [t_i, t_{i+1})$, then $VR_t(d) = VR'_t(d)$. Therefore it is possible to evaluate $\{VR_t(d)\}_{t\geq 0}$ using a finite number of computations.
We can calculate the homology groups for $VR_t(d)$ and consequently obtain vector spaces $V(t)$ parametrized by t. The structure of such vector space is very useful to simplify the underlying information in a dataset, as expressed by the following theorem.

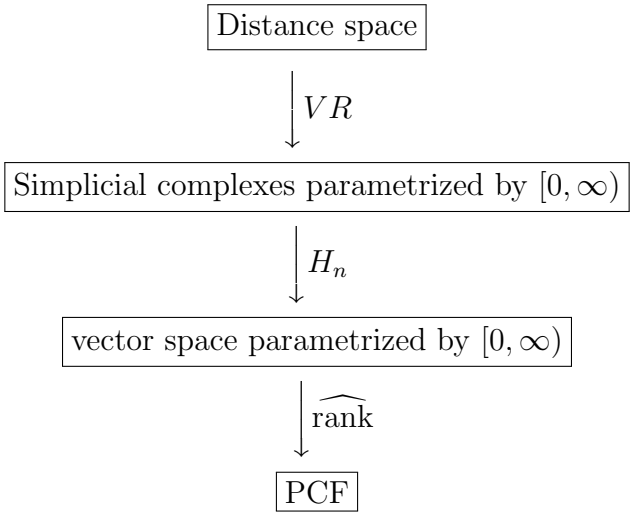
Theorem 2.14. *Any tame vector space parametrised by* $[0, \infty)$ *is isomorphic to a direct sum of bars* $\bigoplus_{i=1}^{r}[a_i, b_i)$.

This follows from the fact that two finite dimensional vector spaces are isomorphic if and only if they have the same dimension. The sequence $\{(a_i < b_i)|i = 1, ..., r\}$ is called the *bar code* of V.

Finally, we can extract a piecewise constant function from this direct sum of bar codes, called the stable rank.

Definition 2.15. The *stable rank* of a tame parametrised vector space V which has a bar decomposition according to 2.14 is given by $\widehat{\text{rank}}\, V(t) = \{\#i \mid b_i - a_i \geq t\}$.

In summary, the TDA pipeline used in this project is given by:



2.1.8 Stability of the Process

In section 2.1.7 we defined a method to convert a distance space into a PCF. For practical applicability it is of key importance to investigate if this pipeline is stable, that is if two distance spaces are similar in structure, their stable ranks should also be similar.
A way of defining similarity between two pseudometric spaces is through the *Gromov-Hausdorff* (GH) distance. The GH distance features various structural properties that make it particularly useful for this purpose, a more detailed description can be found in [6]. As a similarity measure for piece-wise constant functions the interleaving metric can be used. Using these two measures, it turns out that the process described in section 2.1.7 does indeed fulfil the stability property as expressed by theorem 2.16.

Theorem 2.16. *Let X and Y be two pseudometric spaces. When calculating the corresponding n-th homologies and processing them to stable ranks the following inequality holds.*

$$2\mathrm{GH}(X,Y) \geq d_{\bowtie}\big(\widehat{rank}\, H_n(X), \widehat{rank}\, H_n(Y)\big) \tag{5}$$

2.1.9 Hierarchical Clustering

Hierarchical clustering is a method to map a distance space to a dendrogram. Dendrograms are useful since they provide a clear visual representation of a dataset.
Therefore, we apply the following method. We choose an extension ψ of d and we construct a sequence of partitions of X, $P_0 > P_1 > ... > P_n$ by merging two blocks p, q of P_i in P_{i+1} if $\psi(p,q) \leq s$. At the end of this process we obtain a dendrogram. It turns out that such dendrograms are well-related to the general process.

Proposition 2.17. *Let K be a field, and for a set X we define $KX = \bigoplus_{x\in X} K$. Then, KX/d_t coincides with $H_0(VR_t(d), K)$.*

This means that computing H_0 of a distance space reveals the same information as computing the dendrogram from hierarchical clustering.

2.2 Support Vector Machines

2.2.1 Primal Problem

Let $\mathcal{D} = (x_t, y_t)_{t=1}^n$ be a data set with data points x_t in $\mathbb{R}^k$ and labels y_t in $\mathbb{R}$. Support vector machines are a classification scheme whose main objective is to linearly separate $\mathcal{D}$ set into two classes. Linear classifier are generally given by

$$\mathcal{F} = \{f : f(x) = sign(x^T\theta) \text{ for some } \theta \in \mathbb{R}^k\} \tag{6}$$

A linear classifier is more robust the larger its geometric margin $\gamma_{geom} = \frac{\gamma}{\|\theta\|}$ is, where

$$\gamma = \min_{t=1,...,n} y_t x^T\theta \tag{7}$$

Since we are interested in finding robust classification algorithms, the goal is to maximize this margin.

$$\max_{\theta,\gamma} \gamma_{geom} \text{ subject to } y_t x^T\theta \geq \gamma \ \ \forall t \in \{1,...,n\} \tag{8}$$

This can be rewritten as a minimization problem that is more convenient to solve in practice.

$$\min_{\tilde{\theta}} \frac{1}{2}\|\tilde{\theta}\|^2 \text{ subject to } y_t x^T\tilde{\theta} \geq 1 \ \ \forall t \in \{1,...,n\} \tag{9}$$

where $\tilde{\theta} = \frac{\theta}{\gamma}$. Including an offset or bias θ_0 one can generalize the class of linear classifiers.

$$\mathcal{F} = \{f : f(x) = sign(x^T\theta + \theta_0) \text{ for some } \theta \in \mathbb{R}^k, \theta_0 \in \mathbb{R}\} \tag{10}$$

Using this as basis for support vector machines, the *primal problem* [7] is given by

$$\min_{\theta,\theta_0} \frac{1}{2}\|\tilde{\theta}\|^2 \text{ subject to } y_t(x^T\theta + \theta_0) \geq 1 \ \ \forall t \in \{1,...,n\} \tag{11}$$

This problem has in fact a unique solution if the data set can be separated by a hyperplane. However, most data sets are not linearly separable. Therefore, it us useful to allow for misclassified examples but to penalize them. This formulation is called the *soft-margin SVM.*

$$\min_{\theta,\theta_0,\zeta} \frac{1}{2}\|\tilde{\theta}\|^2 \text{ subject to } y_t(x^T\theta + \theta_0) \geq 1 - \zeta_t \ \ \forall t \in \{1, ..., n\} \tag{12}$$

$\zeta = (\zeta_1, ..., \zeta_n)$ denotes the so called regularization parameter. In this formulation SVMs belong to the class of convex optimization problems which can be addressed efficiently by numerical general-purpose solvers.

2.2.2 Kernel Trick

The previous definition of support vector machines is applicable for data points in $\mathbb{R}^k$ and the classification is based on the standard scalar product. However, this procedure can be further extended. The idea is to introduce a *kernel function* $k : \mathcal{X} \times \mathcal{X} \to \mathcal{V}$ that maps from the input space $\mathcal{X}$ to a space $\mathcal{V}$ with an attached inner product

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{V}} \tag{13}$$

where $\phi : \mathcal{X} \to \mathcal{V}$ is the so called *feature map.* A clever choice of ϕ ensures that we can efficiently calculate the value of the kernel function without explicitly mapping to $\mathcal{V}$. Sums and products of kernel functions again form kernel functions.
Replacing the inner product in equation 11 with a kernel mapping to $\mathbb{R}$, the primal problem for kernelised support vector machines [7] is given by

$$\min_{\theta,\theta_0} \frac{1}{2}k(\theta, \theta) \text{ subject to } y_t(k(x, \theta) + \theta_0) \geq 1 \ \ \forall t \in \{1, ..., n\} \tag{14}$$

Given the framework of H_n stable ranks, kernel based methods are a particularly suited tool in conjunction with topological data analysis. The kernel that we use for our implementation of SVMs is the L^2 inner product given the stable ranks $\widehat{\text{rank}}_n X(t)$ for the n-th homology [1].

$$k_n(X, Y) = \int_0^\infty \widehat{\text{rank}}_n X(t) \, \widehat{\text{rank}}_n Y(t) dt \tag{15}$$

3 Methods

3.1 Dataset

The Facial Expression Recognition 2013 (FER2013) data set is composed of 35685 examples of 48×48 grayscale images from seven different emotions. Those images are classified into seven categories based on the emotion shown in the facial expressions: *happiness, neutral, sadness, anger, surprise, disgust, fear*. In figure 1 one image per category is displayed.

The data set was created for a workshop challenge on Kaggle. The images were collected by Pierre Luc Carrier and Aaron Courville from the University of Montréal [2]. They were obtained and labelled through the Google Search API using agglomerations of terms that are related to one emotion such as "blissful" for being happy and "enraged" for being angry. These keywords were combined with expressions that ensure an equal distribution of genders, ethnics, ages and other factors amongst the images. The first 1000 images that were returned were manually inspected in order to correct erroneous labels and remove duplicates. OpenCV face recognition was used to detect the silhouette of each face and crop the images to the final 48×48 gray scale format.
Consequently, the FER2013 data is a well pre-processed set that remains amongst the standards for FER benchmarking [3, 4, 5]. Nevertheless, with around 68% human accuracy it also stresses the difficulty of this task.

3.2 Topology on the Level of Emotions

In order to perform topological analysis on the level of emotions we represent the data in a particular way that is denoted by embedding 1. Therefore, the set is divided into seven categories according to the emotions. Each category comprises of a point cloud and the data points in such a cloud are arrays of length 48×48 that contains integer values between 0 and 255 representing the grayscale values of the pixels of an image. Since the categories are composed of different numbers of images and potentially contain outlier, we employ subsampling. We randomly select 100 picture per emotion and repeat this process 25 times. This increases the robustness of topological analysis with respect to the selection of the subset.
Based on this embedding, we calculated the H_0 and H_1 stable ranks of the set. For

Figure 1: The FER2013 data set comprises of images from 7 different emotions. The illustration exemplified each category by one picture.

H_0 we apply the cityblock distance together with the single linkage extension since this combination best elucidates the difference between the different categories. The H_1 stable ranks are computed using the l_p metric with metric parameter $p = 3$ since that again resulted in the clearest distinction. Furthermore, we perform hierarchical clustering and compute the barcodes with the same distance and linkage as for the stable ranks to put the results into perspective.

3.3 Topology on the Level of Images

In order to perform topological analysis on the level of pictures we use a second embedding. Each image is represented as a point cloud of 48×48 points of the form (x_i, y_i, z_i) where x_i and y_i are the coordinate values of a pixel that range between 1 and 48 and z_i is its grayscale value between 0 and 255. An illustration to explain this process is displayed in Figure 2.

Figure 2: For embedding 2 we represent each image as a point cloud. Each datapoint corresponds to a pixel. The red pixel is stored as $(1, 1, 24)$, the yellow pixel as $(10, 24, 29)$ and the blue pixel as $(48, 48, 195)$. After normalization the representations are given by $(0, 0, 9.41)$, $(21.28, 50, 11.37)$ and $(100, 100, 76.47)$.

We again use subsampling to obtain 200 images from each emotion category and normalize the (x_i, y_i, z_i) coordinate values such that they lie between 0 and 100 in order to increase the influence of the position compared to the grayscale values. The number of images that we could select was limited by the computational cost of computing the homologies in this case. The H_0 and H_1 stable ranks are determined using single linkage clustering and the l_p metric with metric parameter $p = 2$ since this yielded the clearest distinction. Using the interleaving metric, we calculate the average stable ranks and the standard deviation of the stable ranks per emotions to evaluate how similar the topologies of the images within one category are.
We further analyse the separability by constructing support vector machine as described in section 2.2. This allows us to test the pairwise linear separability of the different emotions in the space of PCFs. In addition to the canonical kernels $k_0(X, Y)$ and $k_1(X, Y)$ we also apply their sum as kernel.

$$k_{0+1}(X, Y) = k_0(X, Y) + k_1(X, Y) \tag{16}$$

The implementation follows the idea from Agerberg et al. [1]. Our code is available on github [8]. We split the data into 80% training and 20% test and subsequently perform 5-fold cross validation to receive the average accuracy. We compare the results of the SVMs using the topological kernels to standard SVMs that use the euclidean inner product. In the latter case each image is given as a 48×48 vector with values between 0 and 255 for each pixel.

4 Results

4.1 Embedding 1

The H_0 stable ranks that were computed using the cityblock distance together with the single linkage extension are displayed in figure 3. As described in the methods section we make this choice since it gives the most meaningful results to in the sense that the stable ranks of different emotions are visually distinguishable. We indeed note the emotion "disgust" is clearly set apart from the others. A possible explanation for this observation could be that the facial expressions related to disgust are the most contorted ones. The use of the different metrics yields similar results although the stable ranks do not separate as well.
The H_1 stable ranks are shown in figure 3 using the euclidean metric. In this case, the emotion *surprise* is minorly set apart from the other emotions.
The dendrogram that links each emotions is represented figure 4. It was created using 400 randomly selected pictures from each emotion category and average linkage was used. We observe that "disgust" and "happy" cluster together rapidly, which is a bit unexpected regarding to the previous results using H_0 stable rank where "disgust" was the emotion that clearly separated from the others.
Since the cityblock metric was used in order to compute stable ranks, figure 7 in the supplementary displays the bar codes. It can be seen that the bar code of "disgust" differs slightly from the other categories when considering H_0 homology.

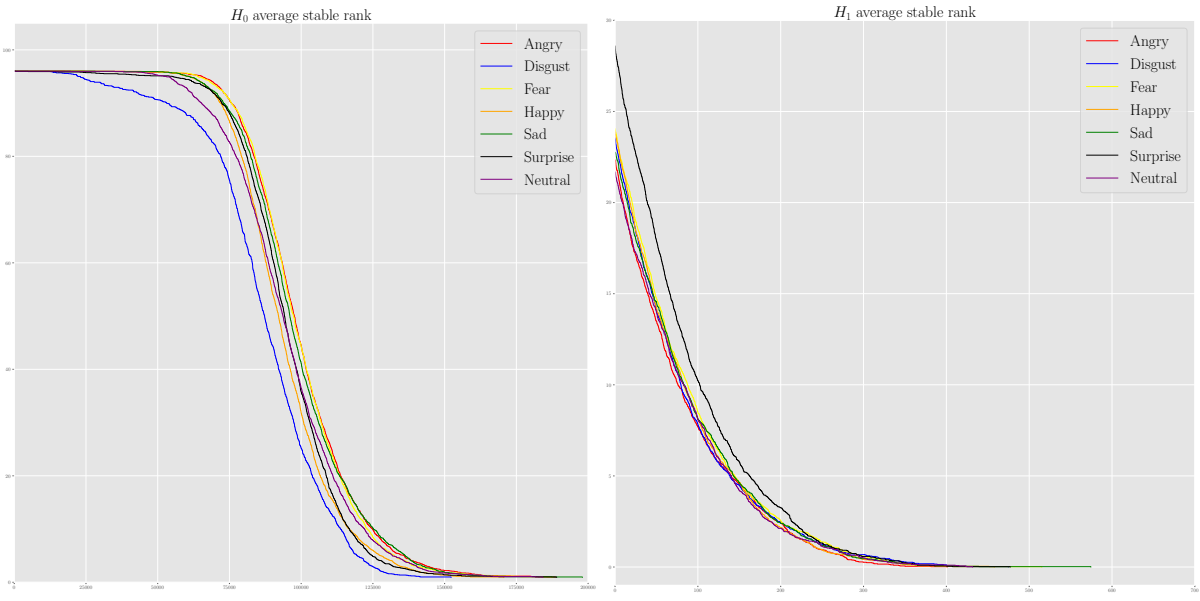


Figure 3: The average H_0 and H_1 stable ranks for the different emotions were calculated through subsampling 100 images and repeating this process 25 times.

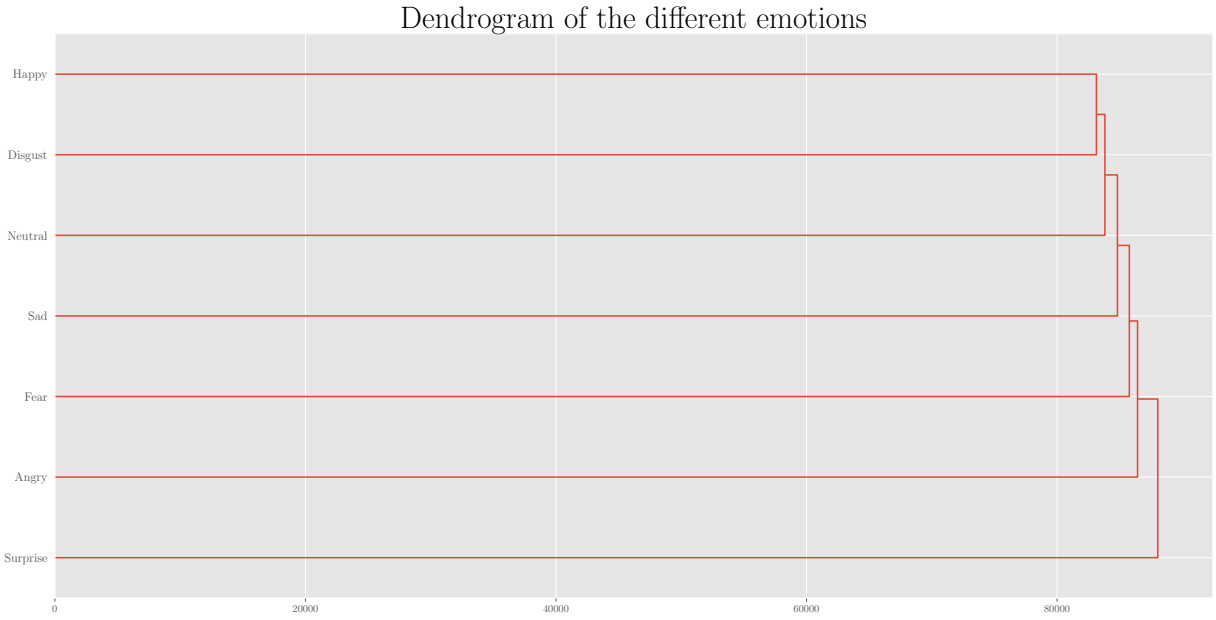


Figure 4: The above dendrogram was created using the cityblock metric together with the average linkage extension on embedding 1.

4.2 Embedding 2

The H_0 and H_1 average stable ranks for the second embedding are displayed figure 5. All stable ranks from each category are shown in figures 8 and 9 in the supplementary. Since the results are visually almost indistinguishable, we further computed the standard deviation for each emotion category using the interleaving distance. The results are shown in table 1 and 2.
The differences in average stable ranks for different emotion categories are difficult to identify. Moreover, the emotion categories "disgust" and "fear" have the lowest standard deviation for H_0 while "fear" and "happy" feature the lowest standard deviation for H_1. Ideally, low standard deviation implies high similarity of images from the same category despite originating from different people. This judgement is still in doubt since our topology may extract noises rather than emotion information encoded in the images.

Emotion Category	Standard Deviation
Angry	19.4792
Disgust	15.8297
Fear	15.2311
Happy	25.3311
Sad	24.5841
Surprise	18.8891
Neutral	23.1706

Table 1: The standard deviation of H_0 stable ranks for embedding 2 was calculated using the interleaving distance.

Emotion Category	Standard Deviation
Angry	10.6946
Disgust	14.6055
Fear	8.5222
Happy	9.4151
Sad	11.4346
Surprise	14.2093
Neutral	14.7785

Table 2: The standard deviation of H_1 stable ranks for embedding 2 was computed using the interleaving distance as well.

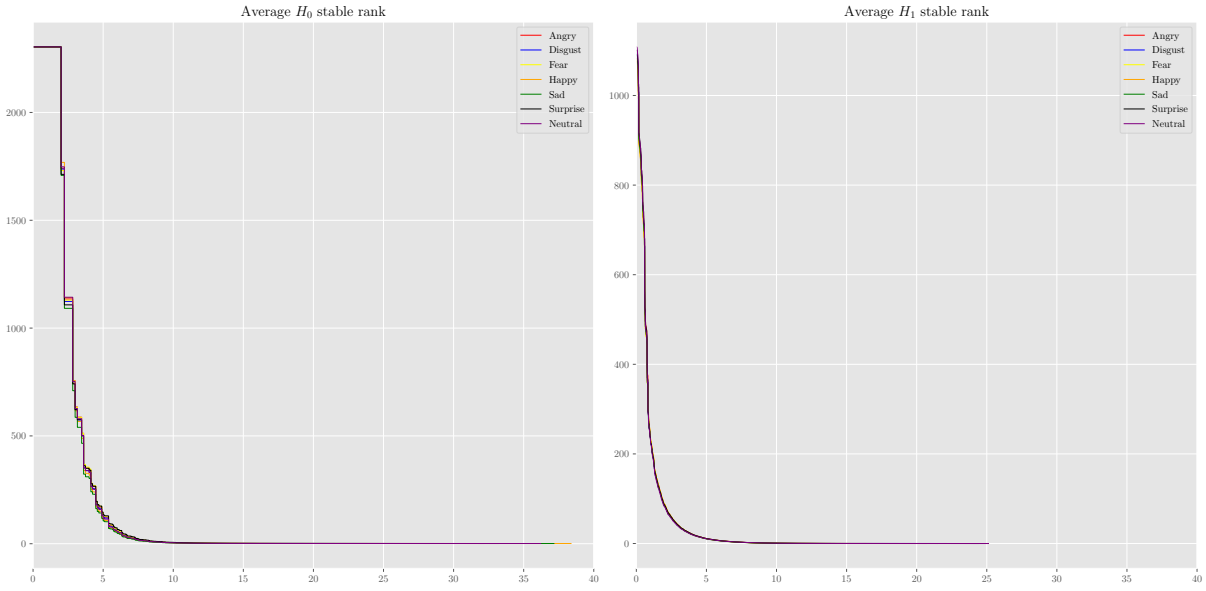


Figure 5: The average H_0 and H_1 stable ranks for embedding 2 were determined through subsampling 200 images from each category.

4.3 Classification Using Support Vector Machines

As noted in the previous section, it is hard to come to a conclusion regarding the classifiability of the seven emotions purely from a visual inspection of the stable ranks in embedding 2. Using the kernelised version of support vector machines that was introduced in section 2.2 it is possible to specify this observation by characterising distinguishability as linear separability. The accuracy of SVMs using the standard inner product, the k_0, the k_1 as well as the k_{0+1} kernels are displayed in figure 6.
Standard SVMs achieve moderate performance that is comparable to a random predictor. The k_0 and k_1 kernelised versions display similar performance. The summarized kernel k_{0+1} results in slightly better accuracy. In addition, their is no clear tendency for certain pairs of emotions to be easier separable than others. Thus, we conclude that is the application of SVMs provides support to the hypothesis that it remains difficult to classify human facial expressions using the tools from TDA presented in section 2 applied to the second embedding.

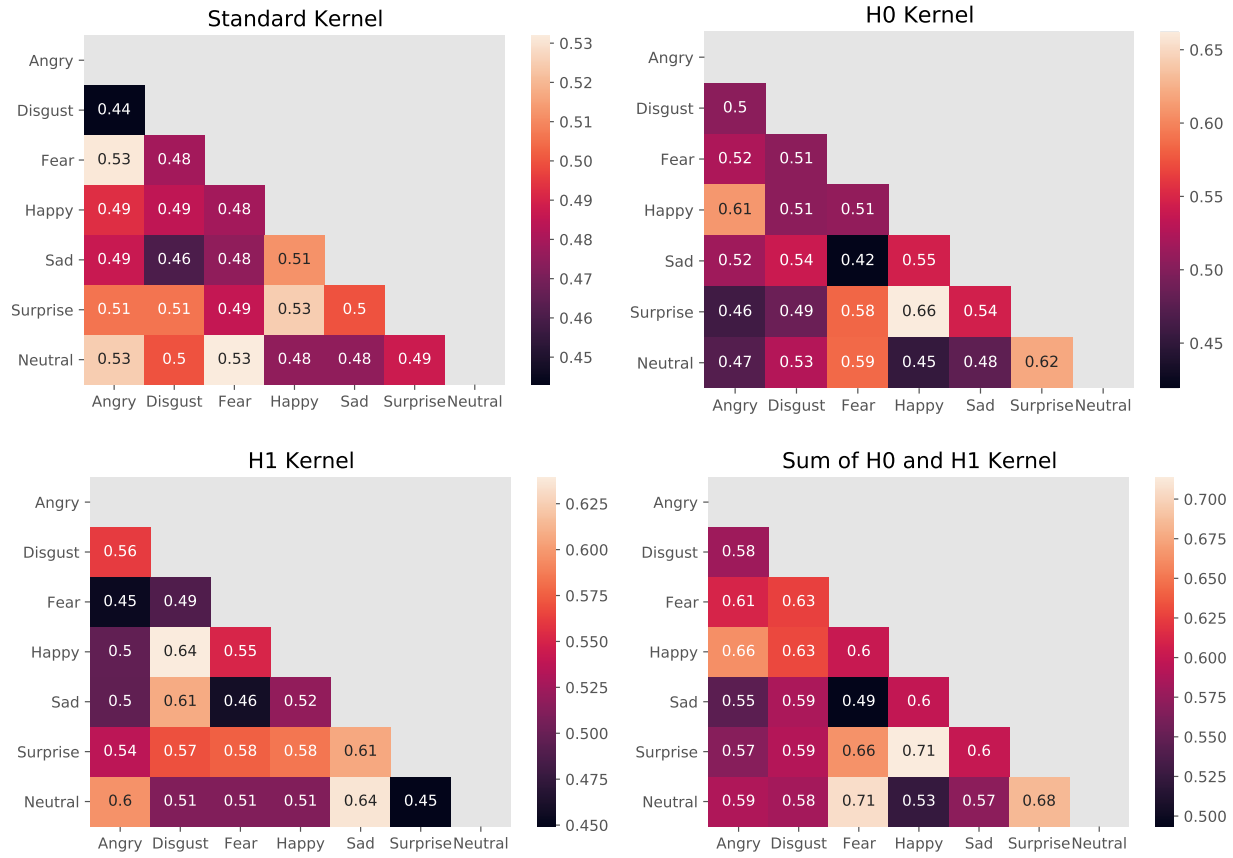


Figure 6: The accuracy of the SVMs using the topological kernel k_0 was determined using 5-fold cross validation.

5 Conclusion

Facial expression recognition under realistic circumstances is a long-standing problem. The results in our project indicate that tools from topological data analysis struggle to overcome the main difficulties related to this task as well.
Nevertheless, we could observe some successful characterisations. In the first embedding used to analyse topology on the level of emotions, the H_0 stable ranks of "disgust" clearly separate from the other categories. A possible explanation for this observation could be that "disgust" is linked to stronger mimics. When analysing the results from embedding 2, which captures topology on the level of images, we found it much harder to formulate clear findings. Both the H_0 and H_1 stable ranks resulting from the images in this representation are visually almost indistinguishable regardless of the underlying facial expression. This can be further specified by assessing the linear separability using kernelised support vector machines on H_0 and H_1 stable ranks. Although we could improve the performance by combing the H_0 and H_1 kernels together, the classification accuracy still remains close to a random predictor.
In general, there are certain limitations and drawbacks to the approach that we took in this project which could be the reason for the observed outcomes. Due to the related computational costs we had to refrain from computing H_2 or higher homologies. For similar reasons we only subsample 200 images from each category in embedding 2. Since the entire training data set features a larger variety of images this somewhat limits the generalizability of our results. Other state-of-art machine learning methods such as convolutional neural networks could make use of all training images [4, 5]. Consequently, it remains in doubt if the moderate performance originates from inconsistencies in the data set, the lack of topological features or SVMs being a too simple classifier. Nonetheless, we see that it was possible to slightly improve the accuracy of standard SVMs by applying TDA based kernels. Therefore, other more involved machine learning algorithms could potentially benefit from the inclusion of topological features as well.

6 Supplementary Figures

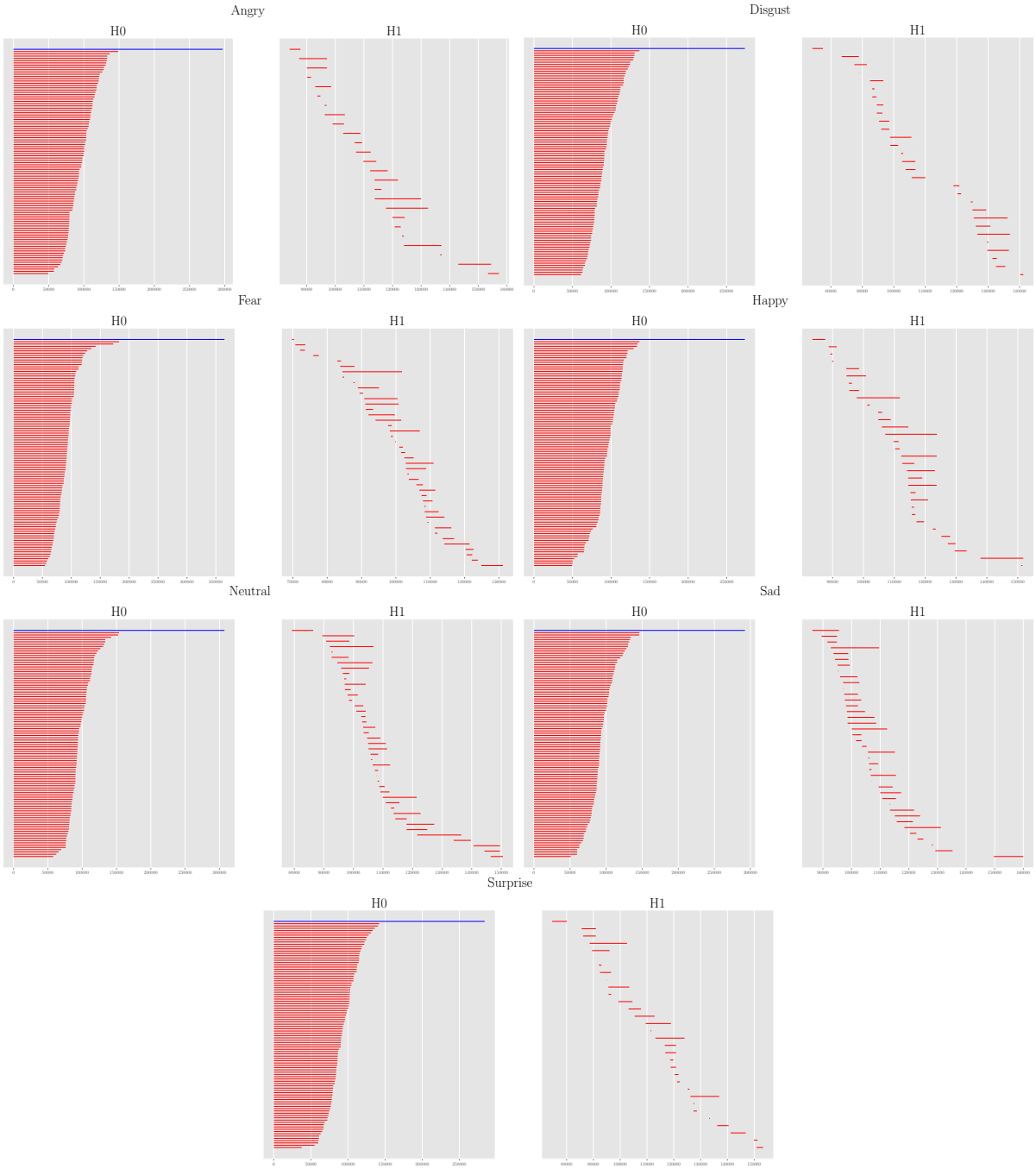


Figure 7: The above shown bar codes of H_0 and H_1 for the different emotions were computed using the cityblock metric.

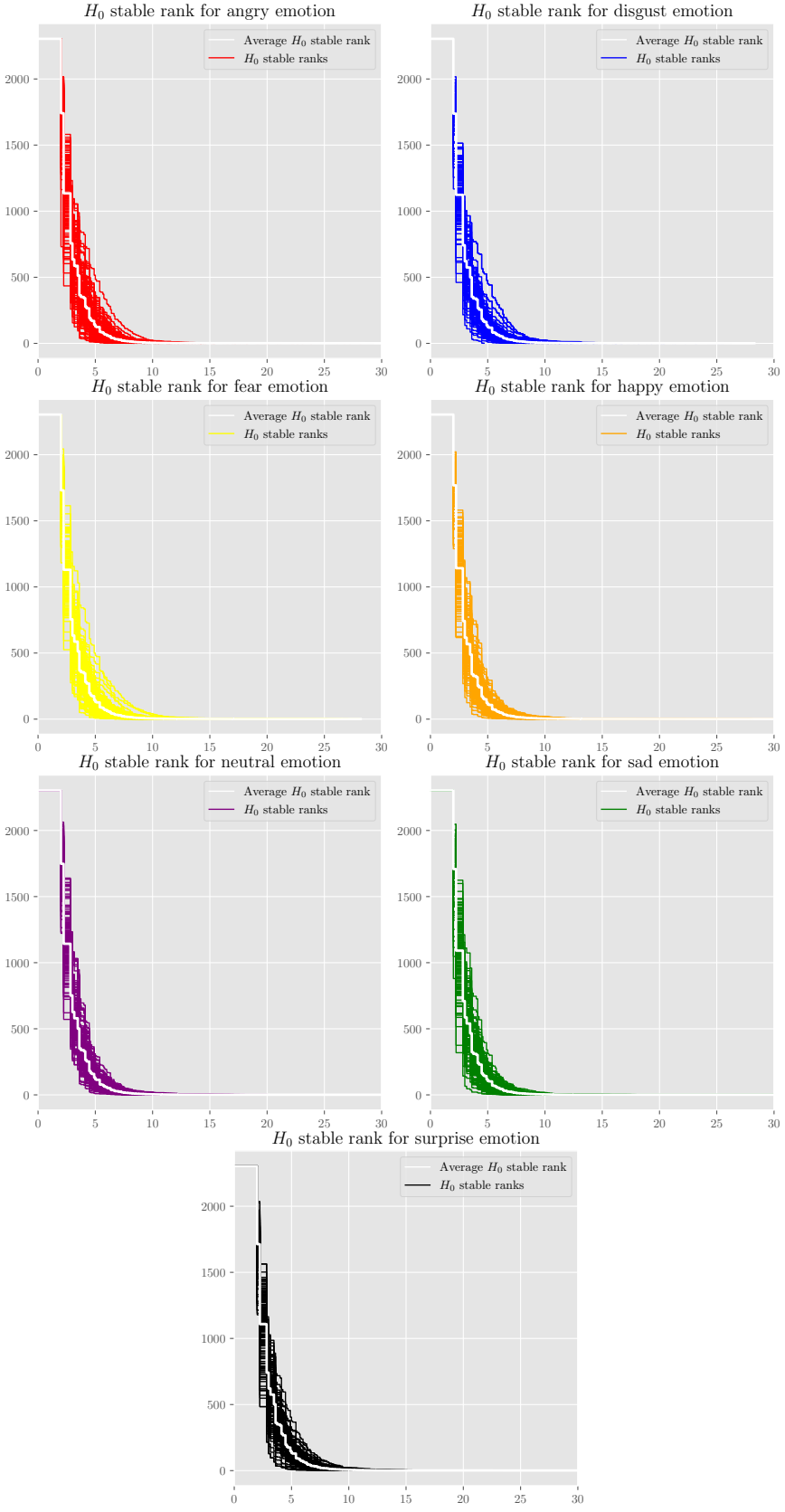


Figure 8: For the images above we calculated the H_0 stable ranks for 200 images from each category and display them together with the respective average.

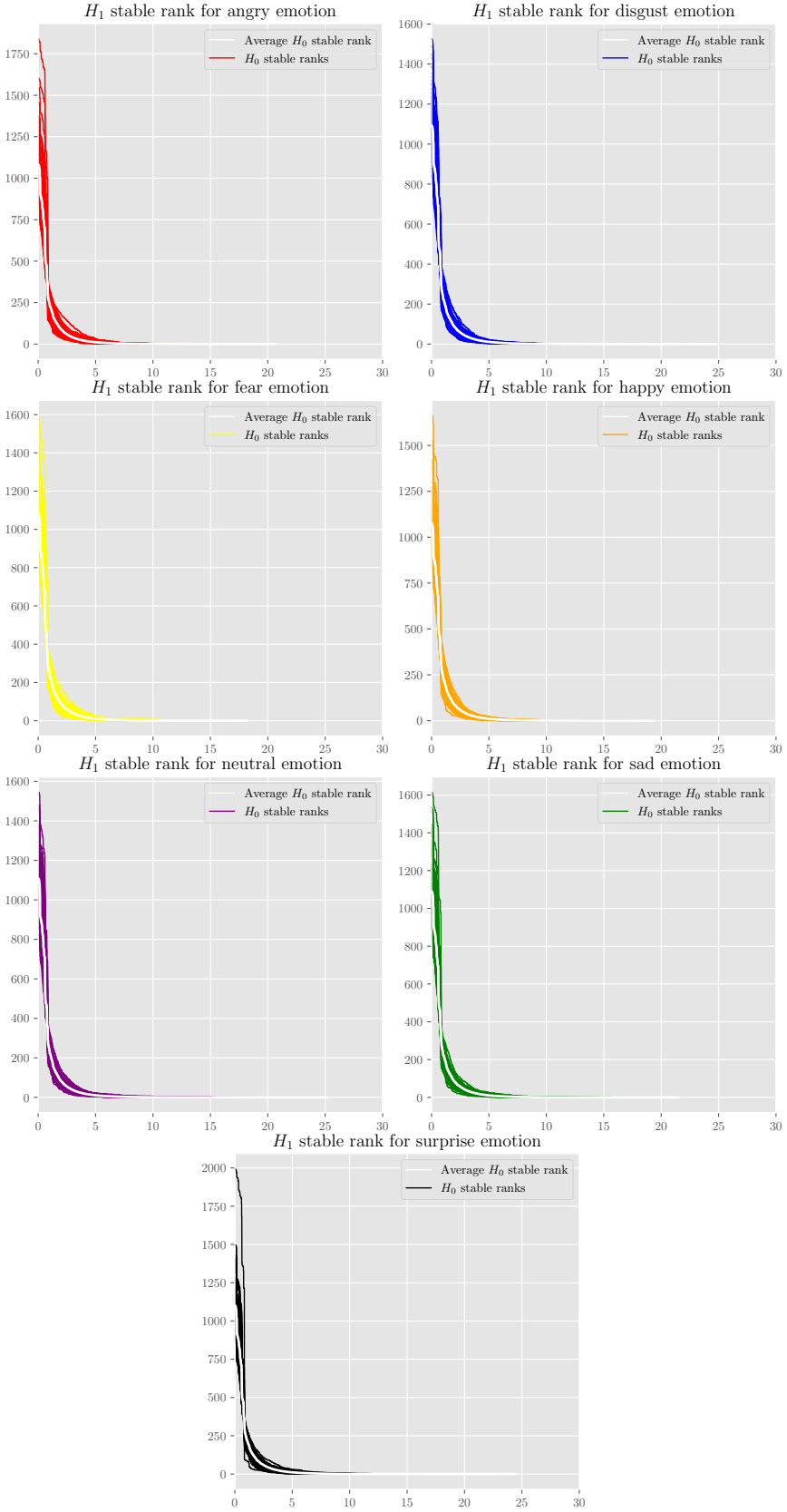


Figure 9: Similar to figure 8 we proceeded to calculate 200 H_1 stable ranks per emotion and the corresponding averages.

References

[1] Jens Agerberg, Ryan Ramanujam, Martina Scolamiero and Wojciech Chachólski (2021) Supervised Learning Using Homology Stable Rank Kernels, Frontiers in Applied Mathematics and Statistics, Volume 7

[2] Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, Yoshua Bengio (2015) Challenges in representation learning: A report on three machine learning contests, Neural Networks, Volume 64, Pages 59-63

[3] Mahmood, Awais and Hussain, Shariq and Iqbal, Khalid and Elkilani, Wail S. (2019) Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion, Mathematical Problems in Engineering 9185481

[4] Yousif Khaireddin and Zhuofa Chen, Facial Emotion Recognition: State of the Art Performance on FER2013, CoRR 2105.03588

[5] S. Minaee, M. Minaei, A. Abdolrashidi (2021) Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, Sensors 21,3046

[6] Wojciech Chachólski (2021) Topological data analysis Lecture notes

[7] Christopher M. Bishop (2006) Pattern Recognition and Machine Learning (Information Science and Statistics, 2nd Edition, Springer-Verlag

[8] Yuqi Shao, Lea Keller, Maxime Scali, Niclas Popp, 2020, https://github.com/yuqish/TDA_project