

# Feature Clock: High-Dimensional Effects in Two-Dimensional Plots

Olga Ovcharenko\*  
ETH Zürich

Rita Sevastjanova†  
ETH Zürich

Valentina Boeva†  
ETH Zürich

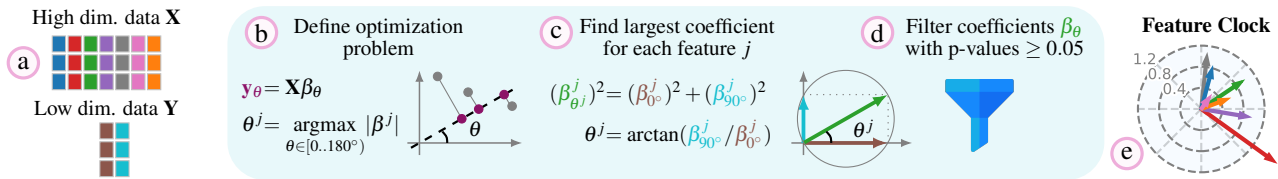


Figure 1: Feature Clock uses **a** high  $\mathbf{X}$  and low  $\mathbf{Y}$  dimensional data. **b** We define the feature contribution as a coefficient of the linear regression (LR) between  $\mathbf{X}$  and  $\mathbf{y}_\theta$  ( $\mathbf{Y}$  projected at angle  $\theta$ ). The optimization goal is to find the angle  $\theta^j$  at which the coefficient  $\beta_{\theta^j}^j$  of feature  $j$  is maximized. **c** The solution is derived from LR coefficients of  $\theta = 0^\circ$  and  $\theta = 90^\circ$ . **d** We filter the insignificant LR coefficients based on p-values. Features with insignificant p-values are not visualized. **e** The largest coefficient of each feature is visualized in the Feature Clock.

## ABSTRACT

Humans struggle to perceive and interpret high-dimensional data. Therefore, high-dimensional data are often projected into two dimensions for visualization. Many applications benefit from complex nonlinear dimensionality reduction techniques, but the effects of individual high-dimensional features are hard to explain in the two-dimensional space. Most visualization solutions use multiple two-dimensional plots, each showing the effect of one high-dimensional feature in two dimensions; this approach creates a need for a visual inspection of  $k$  plots for a  $k$ -dimensional input space. Our solution, Feature Clock, provides a novel approach that eliminates the need to inspect these  $k$  plots to grasp the influence of original features on the data structure depicted in two dimensions. Feature Clock enhances the explainability and compactness of visualizations of embedded data and is available in an open-source Python library<sup>1</sup>.

**Index Terms:** High-dimensional data, nonlinear dimensionality reduction, feature importance, visualization.

## 1 INTRODUCTION

Dimensionality reduction methods transform high-dimensional data into lower-dimensional space. These methods aim to preserve various properties of the original data in lower dimensions (*e.g.*, variance, pair-wise distances or similarities, or grouping structure). Use cases include numerous applications: feature selection [18, 36], visualization [17, 34, 26], compression [35, 13], and approximate techniques to avoid the curse of dimensionality [12, 1]. There are two main types of dimensionality reduction: linear and nonlinear.

Linear dimensionality reduction (LDR) techniques linearly project higher dimensional data into a lower dimensional space. All LDR methods can be seen as a single matrix multiplication, according to  $\mathbf{Y} = \mathbf{X}\mathbf{W}$  where  $\mathbf{X}$  is the original data with samples as rows and features as columns,  $\mathbf{Y}$  is the low-dimensional representation, and  $\mathbf{W}$  is the transformation matrix. One of the most common visualization techniques to show the effect of high-dimensional features in LDR space is a biplot [10], which depicts the rows of  $\mathbf{W}$ .

Nonlinear dimensionality reduction (NLDR), also called manifold learning, is a set of techniques that aim to project high-dimensional data onto lower-dimensional manifolds [32, 23, 24]. NLDR algorithms are usually better at preserving data grouping structure in two dimensions than LDR; this is a desirable behavior found in the outputs of many NLDR methods [32, 33, 23, 27, 24]. A classical example of NLDR is t-distributed stochastic neighbor embedding (t-SNE) [32, 33], which constructs two probability distributions over

high- and low-dimensional data and minimizes the Kullback–Leibler divergence (KLD) [16] between the two distributions. A disadvantage of NLDR is that the distances in the low-dimensional space may not reflect the distances in the original space [19]. Constructing biplots for NLDR is impossible because no linear  $\mathbf{W}$  exists to explain the effect of features. Currently, one can visualize the effect of each feature of interest in a separate plot [34, 17, 21]. These numerous plots are one of the best methods to understand the original feature’s effects in low-dimensional spaces, but this solution is not scalable.

This paper introduces three types of static visualizations, highlighting the contributions of the high-dimensional features to linear directions of the two-dimensional (2D) spaces produced by NLDR. The three techniques are a Global Feature Clock indicating the direction of features’ contributions in low-dimensional space for the whole dataset, a Local Clock explaining features’ impact within selected points, and an Inter-group Clock visualizing contributions between groups of points. The implementation is an open-source Python package. Our technical contributions include: (1) Feature Clock, a novel technique for plotting feature contributions using linear regression in Sec. 3, (2) a formal proof that ensures a correct behavior of Feature Clock in Sec. 3 and Suppl. Materials, and (3) an experimental evaluation of the proposed visualization technique in several application cases in Sec. 4.

## 2 BACKGROUND AND RELATED WORK

This section summarizes common approaches to visualize the results of LDR and NLDR methods and highlights their limitations.

**Biplot for LDR:** Biplot [10] is a visualization technique that can be applied to all LDR techniques. It involves creating a scatter plot that represents the low-dimensional data points, called a *score plot*. Vectors are depicted to show the strength of each feature influence (rows of  $\mathbf{W}$ ), which are referred to as a *loading plot*. By interpreting a biplot, a user extracts information about the direction and strength of the association between original and low-dimensional space features. The effect of  $\mathbf{W}$  is uniform across the low-dimensional space, making biplot an effective tool. Fig. 2a shows a PCA biplot for the Iris flower dataset [8] where the respective x- and y-axis are the two principal components. The points are 2D embeddings of the four-dimensional data points representing realizations of the iris plants. The arrows point toward change for a given feature. For instance, sepal width points at  $\sim 100^\circ$ , meaning that increasing sepal width translates to moving points up at the same angle. A change in petal length and width affects embedded coordinates in the same direction. The magnitude of an arrow indicates how significant a change of the original feature affects shifts in coordinates.

**Visualizations of feature effects for NLDR:** While NLDR is popular for data visualization, interpreting the global and local struc-

\*e-mail: oovcharenko@ethz.ch

†e-mail: {rita.sevastjanova|valentina.boeva}@inf.ethz.ch

<sup>1</sup><https://pypi.org/project/feature-clock>

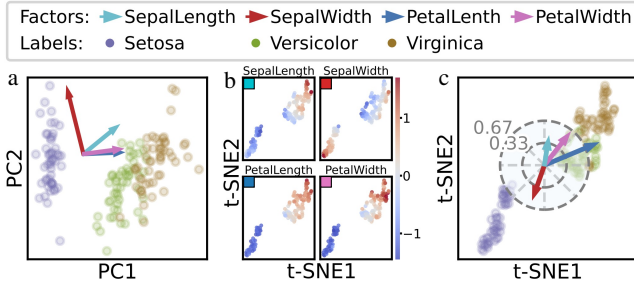


Figure 2: (a) PCA biplot with feature loadings, (b) t-SNE scatter plots for each high-dimensional feature (perplexity=17). (c) A Feature Clock showing the feature contributions for the Iris dataset.

tures of NLDR outputs can be challenging. A common method to assess the retained information in the low-dimensional representation involves assigning color codes to each projected data point based on a feature from the input data [20], which often results in visual clutter. Thus, field-based techniques can be used to reconstruct a 2D scalar function for a feature of interest used for spatial color coding [29]. Other approaches illustrate regions of maximum attribute values. For example, DataContextMap [4] enhances the projection by adding extra data points representing high feature values and displaying contours to depict feature distribution. DimReader [7] uses perturbations, measures their effect on the projection’s outcome, and visualizes the resulting changes as grid lines. t-viSNE [2] supports exploring t-SNE projections, employing visualizations to analyze correlations between dimensions and visual patterns.

Many methods aim to identify clusters in the 2D space and explore their properties in the high-dimensional data. For instance, Joia et al. [14] propose a visualization that utilizes matrix decomposition to define clusters in the visual space. Turkyay et al. [31] associate representative factors with groups of dimensions for iterative analysis of the projected data, presenting techniques to compare and evaluate these factors. Several approaches focus on comparing the properties of different groups found in the 2D space. For example, Probing Projections [30] enable to examine projections based on approximation errors and the influence of data dimensions on the projection space, and supports the comparison of different clusters in the 2D space. Eckelt et al. [6] introduce new designs for visual summaries of common properties of such clusters and propose novel group comparison visualizations. Marcilio-Jr and Eler [22] employ cluster-oriented analysis to explain dimensionality reduction techniques with the so-called ClusterShapley method. Fujiwara et al. [9] present a visual analytics method that explains features’ contributions to the projected data by calculating each feature’s relative contribution to the contrast between one cluster and other clusters.

Despite the work on how to visualize NLDR embeddings, there is no biplot alternative for NLDR. A common solution is to show a plot per feature to analyze high-dimensional values in the low-dimensional space, see Fig. 2b. Multiple scatter plots are a space-consuming and not scalable solution. We suggest an alternative to the biplot: Feature Clock shown in Fig. 2c and described in Sec. 3.

### 3 METHODOLOGY

This section outlines the internals of the Feature Clock, see Fig. 1. The Feature Clock helps explore feature contributions of data points projected in two dimensions. A user can be interested in how high-dimensional features contribute to the positions of low-dimensional data points globally or locally, or which features discriminate user-defined classes or clusters. To address the former, we provide the Global and Local Clocks where the arrows correspond to the largest coefficients of linear regression, predicting coordinates of data points projected on a line passing through the center(s) of mass. For the latter, we offer the Inter-group Clock, where arrows correspond to the feature contributions in the classification task (logistic regression).

**Feature Clock Overview:** The method addresses a linear regression problem with the high-dimensional data as the predictor variable,  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , and the low-dimensional data,  $\mathbf{Y} \in \mathbb{R}^{n \times 2}$ , projected on a line at angle  $\theta$  as the target variable,  $y_\theta \in \mathbb{R}^n$  (Fig. 3). We solve the optimization problem  $\theta^j = \operatorname{argmax}_\theta |\beta^j|$  for each feature  $j$ , by finding the angle  $\theta^j$ , at which the absolute value of linear regression coefficient  $\beta_\theta^j$  is maximized. For each feature  $j$ , a closed-form solution allows to find the largest coefficient using coefficients  $\beta_{0^\circ}^j$ ,  $\beta_{90^\circ}^j$ , and Pythagoras theorem (Fig. 1c). First, the high- and low-dimensional data are optionally normalized (default: standardizing high-dimensional input, centering low-dimensional coordinates). Second, the angle and magnitude of the strongest contribution are derived from the linear regression coefficients  $\beta_{0^\circ}$  and  $\beta_{90^\circ}$  (a formal proof in Sec. A.1 of Suppl. Materials). We fit two multivariate linear regression models to find the contribution of high-dimensional features  $\mathbf{X}$  to the low-dimensional projections on the  $x$  ( $y_{0^\circ}$ ) and  $y$  ( $y_{90^\circ}$ ) axes in the 2D space. To compute  $\beta_{0^\circ}$  and  $\beta_{90^\circ}$ , one model is fitted between  $X$  and  $y_{0^\circ}$ , another between  $X$  and  $y_{90^\circ}$ . We advise the user to standardize  $\mathbf{X}$  to make coefficients  $\beta^j$  comparable across features; alternatively, we offer an option to standardize  $\beta^j$ , as *post-processing*. For each feature  $j$ , we define the biggest contribution as  $(\beta_{\theta^j}^j)^2 = (\beta_{0^\circ}^j)^2 + (\beta_{90^\circ}^j)^2$  at angle  $\theta^j = \arctan(\beta_{90^\circ}^j / \beta_{0^\circ}^j)$ . Finally, we use t-test p-values to filter out statistically non-significant coefficients and visualize only significant contributions. The p-value of a linear regression model checks if there is a significant linear relationship between each feature of  $X$  and  $y_\theta$ . If the p-value is low, the relationship is significant (default: p-value < 0.05).

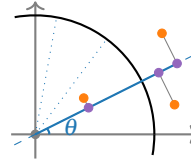


Figure 3: Projection of 2D points on a line at angle  $\theta$ .

An alternative to visualizing one coefficient per variable is to project low-dimensional data on the radius of the semicircle at different angles, see Fig. 3, and visualize all coefficients for all features per projection, see Sec. A.2 of Suppl. Materials. This results in perfect circles that appear because of projection and rotation. Projection angles are defined by increments from  $0^\circ$  to  $180^\circ$  by a user-defined  $\theta = \frac{180^\circ}{m}$  with  $m$  projection lines, where  $\theta = 5^\circ$  is the default value. We default to plot the largest coefficient for each high-dimensional feature since it is a cleaner visualization that maintains maximum information, but the user can opt for the “circles” visualization. If a dataset has many features, a user can visualize only top- $k$  features sorted by impact.

Clock annotations show the coefficient  $\beta_\theta$  size and are intended for an easier comparison within and between clocks. Comparison is possible because all coefficients are normalized. We place the clock in the center of the low-dimensional points, but the user can change the location. The clock location does not influence the computation of the contributions. The size of the clock is defined by the distance between min and max points, but the user can scale the clock.

**Global Feature Clock:** We construct a clock using all data points to visualize the impact of the high-dimensional features in the two-dimensional space. The disadvantage of the Global Clock is that it does not capture finer changes of gradients across the whole data, but rather a trend. Since we approximate the path on a manifold with a line, this clock can miss some information.

**Local Feature Clock:** We use a Local Clock technique to explore data at a finer granularity within a group. The Local Clocks enable an easier analysis of a selection of neighboring points. Class labels or unsupervised clustering of high-dimensional data determine the points used for a single clock. Analyzing original labels gives a perspective on what drives low-dimensional data point coordinates within a particular class or cluster. We default clustering to HDBSCAN with user-chosen parameters that also identifies outliers not assigned to any cluster. Users are responsible for the meaningful

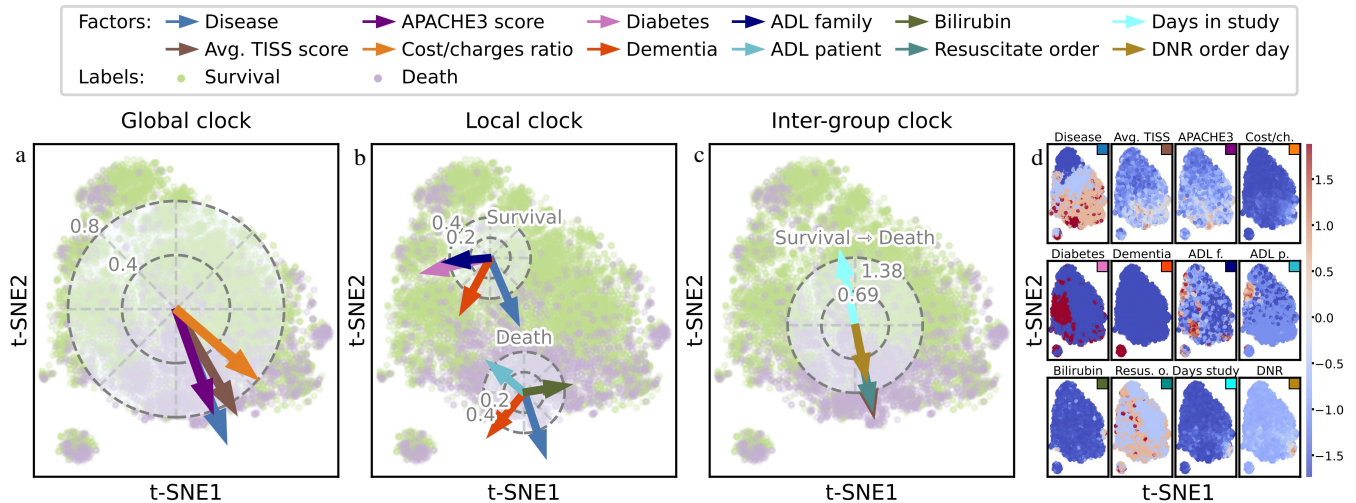


Figure 4: (a) Global Feature Clock shows the general trend for top-4 significant features for hospital survival. (b) Local Clocks help to inspect the data within original labels (patient survived/died in hospital) while (c) Inter-group Clock shows how features change between two classes. (d) t-SNE (perplexity=50) scatter plots support three clocks and show the value range of high-dimensional features for survival in hospital.

clusters and can define their own clusters. For each cluster or class, we apply the same method as in Fig. 1 and create a single clock signifying changes within selected points.

**Inter-group Feature Clock:** The third visualization option helps inspect how variables change between groups, either clusters or user-defined classes. We fit a binary logistic regression with high-dimensional observations as predictor variables and visualize statistically significant coefficients on a single line that connects the group centers. The arrow lengths correspond to the absolute values of the model coefficients. The clock is placed in the center of the line that connects the groups' centers. In a multi-group setting, there is a space limitation for plotting all pairs of groups and their clocks. Therefore, we build a minimum spanning tree (MST) between group centers in low-dimensional space and plot the inter-group clocks only for a trajectory on the MST.

**Limitations:** We explain high-dimensional contributions with linear directions. If data is dense, we can approximate the movement of the gradients on the manifold in the original high-dimensional space with the shortest distance on average in low-dimensional space [3]. We assume that the shortest path is a line. The NLDR maps tend to plot points sparsely located away from cluster centers relatively close to those centers due to the sigmoid shape of the joint probability functions and the form of the objective function (KLD or cross-entropy) [32, 23]. As a result, these points have a minimal impact on the values of the regression coefficients  $\beta^j$ . The Feature Clock technique, especially the Global Clock, can miss information and not always follow the actual manifold and a nonlinear path on it.

**Example:** Fig. 2 shows a PCA biplot with feature loadings, t-SNE scatter plots for each variable, and Feature Clock for the Iris dataset [8]. A biplot (Fig. 2a) shows how features influence low-dimensional representation. We can see that petal width and length are the driving factors for the versicolor and virginica classes. The scatter plots, in Fig. 2b, display how high-dimensional values of each feature are distributed in the 2D t-SNE space. All features except sepal width increase the most with increasing  $x$  and  $y$  values. From a Feature Clock in Fig. 2c, we observe how each variable impacts t-SNE embedding. In the Feature Clock, the sepal length, petal width and length are driving factors for versicolor and virginica classes. Sepal width increases in the direction of the setosa class, similar to the PCA. For PCA, the Feature Clock produces the same arrows as the biplot (figures not shown).

## 4 USE CASES

In this section, we describe several applications of Feature Clocks and how they improve explainability for each task.

### 4.1 Analysis of Low-dimensional Data

Support2 [5] is a dataset that comprises information on critically ill patients across US medical centers. Each patient is assigned one of nine diseases: acute respiratory failure, liver disease, colon/lung cancer, etc. The dataset is often used to train supervised machine learning models to determine whether patients would die in the hospital based on physiological, demographic, and disease information.

There are 46 features describing each patient. It is infeasible to visualize the effects of all features by plotting 46 scatter plots, one for each variable. Therefore, the high-dimensional feature's impact can be visualized by Feature Clocks with the top- $k$  significant features. Fig. 4 shows three clocks and scatter plots supporting clocks for the Support2 data. For all three clocks, we have set  $k = 4$  (top-4 features). We use the hospital death status of the patient as a label for the Local and Inter-group Clocks. We observe that t-SNE captures the difference between patients who die or survive in the hospital.

The Global Clock (Fig. 4a) shows that disease class, cost/charges ratio, average Therapeutic Intervention Scoring System (TISS), and APACHE3 scores differentiate the outcome of critically ill patients. To support the Global Clock, we show scatter plots in Fig. 4d. We see that patients with a less severe disease class survive more often, and medical centers charge these patients less. The average TISS score allows a quantitative comparison of patient care, research experiences of different intensive care units (ICU), and an estimate of the severity of the disease. Most of the deceased patients had a high TISS score. APACHE3 scoring serves as an early warning indication of death and prompts clinicians to upgrade the treatment protocol, making it a useful tool for the clinical prediction of ICU mortality. The people with a negative outcome have a higher APACHE3 score.

The Local Clocks in Fig. 4b represent points that belong to either the survival or death class. We see two clocks; both are driven by the disease class and the presence of dementia. The 2D space of living patients is influenced by variables such as diabetes and the Activities of Daily Living (ADL) score estimated by the family of a sick person. The ADL score evaluates a patient's ability to accomplish their daily activities. The patient's ADL and bilirubin levels are significant contributors to the 2D representation of the patients who die.

In Fig. 4c, the Inter-group Clock shows features that differ between deceased and living patients. As shown by the Inter-group Clock, mortality increases with higher values of the average TISS,

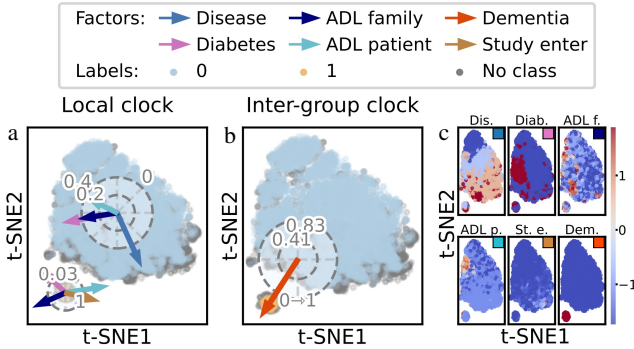


Figure 5: (a) Local Clocks explore the NLDR space by inspecting HDBSCAN clusters. (b) Inter-group Clock shows how features change between clusters. (c) For the verification, we show a t-SNE scatter plot for each feature in the clocks (perplexity=50).

Do Not Resuscitate (DNR) order day, and if resuscitation was given. Survival decreases with days spent in the study.

Instead of visualizing clocks for the original target, the user can explore low-dimensional space with clustering. In Fig. 5, the Local and Inter-group Clocks (top-4 features) use HDBSCAN clusters of high-dimensional data; the resulting two clusters are clearly seen in the t-SNE embedding (blue and orange). Some points are not assigned to any cluster by HDBSCAN (gray). The Local Clock in Fig. 5a shows that coordinates of data points in both clusters 0 and 1 are driven by the presence of diabetes and ADL scores. Interestingly, the ADL score estimated by the patient’s family is high on the left in the two clusters, and contributions point in the respective direction. The same applies to the diabetes status. A bigger cluster 0 is also characterized by disease class while cluster 1 differs by the day of the study entrance.

The Inter-group Clock in Fig. 5b shows the feature that drives the difference between the two clusters. In the t-SNE space, people with dementia are clearly separated from patients without the disease, and that is the only significant feature for the trained classifier.

Further use cases are included in the Supplemental Material.

## 4.2 Neural Network Interpretability

The Pima Indians Diabetes Database [28] provides data on adult females of Pima Indian heritage. It is used to train models predicting if a patient has diabetes based on clinical and demographic data.

The clock technique can be used to explain the hidden states of a neural network (NN). Users can plot a Feature Clock using either a 2D hidden state or apply NLDR to any high-dimensional hidden layer to get a 2D representation. This visualization has the potential to help the user better understand the factors that drive the information extracted at each hidden layer.

In this use case, we analyze the last hidden layer (two neurons) of a NN with two layers. Fig. 6 shows a clock and scatter plots for the last hidden layer, using true labels in the visualization. The

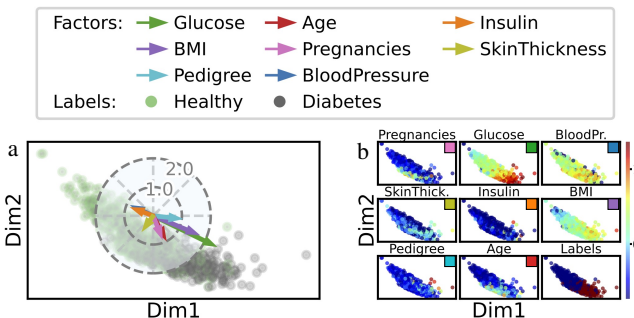


Figure 6: (a) We use Global Clock to describe 2D hidden state of the NN and justify the clock with (b) a scatter plot for each feature.

representations of patients with and without diabetes are almost linearly separable. The clock on Fig. 6 identifies increasing glucose and body mass index (BMI) as indicating factors for diabetes. This conclusion can be verified by scatter plots in Fig. 6, where we show a subplot for each feature, including the target Labels. Age and number of pregnancies also substantially contribute to the activation values of this hidden layer. Low insulin and blood pressure lead to no diabetes, which aligns with low values in the scatter plots.

It is worth noting that all Feature Clock versions are applicable to any NN architecture, including autoencoders. We used the feed-forward NN with two neurons as a simple example.

## 4.3 Preliminary User Feedback

We gathered feedback regarding the readability of the Feature Clock visualization in an informal interview with two post-doctoral researchers from the visual analytics research field. The researchers agreed that the visualization is simple yet effective, easy graspable, and intuitive. They also mentioned potential extensions in highlighting the uncertainty in the feature contribution through the visual blurring of the respective arrows as well as extending the visual representation to those used to represent velocity [11].

## 5 REFLECTIONS, ISSUES AND FUTURE WORK

To summarize, we have shown a group of novel plotting techniques called Feature Clocks. We illustrate how Feature Clocks can be used to interpret lower-dimensional space resulting from NLDR methods and the latent space of deep neural networks.

The implication of the suggested technique is an intuitive, compact, and informative visualization that enhances explainability. It eliminates the need for multiple scatter plots to explain the impact of original features after a non-linear dimensionality reduction. For a large number of variables, making scatter plots for each feature becomes infeasible and challenging to analyze visually. The main limitation is the linear approximation of the nonlinear path on the manifold. This path might not exist if manifolds are disconnected. Feature Clock advises a user about the feature changes, but some information might be missed, especially in the Global Clock, since we might not follow the actual manifold.

Future work may include refinement of the model for the visualization. One can use non-linear regression to visualize complex gradient changes in the feature space. The challenge lies in developing an algorithm that is computationally efficient enough to solve the optimization problem. Another potential direction of future research is replacing  $\mathbf{y}_\theta$  with a trajectory, non-linear in 2D, better reflecting the shortest-distance traverse of the low-dimensional manifold. Additionally, to allow interactive selection of the points to analyze, we could implement a lasso tool using GPU acceleration.

## 6 CONCLUSIONS

We introduced Feature Clock, a group of novel plotting techniques for global, local, and inter-group analysis of non-linear two-dimensional spaces. First, the Feature Clock enables compact representations of the two-dimensional space, highlighting contributions of high-dimensional features. Second, we found an efficient way to identify the largest feature contribution: We mathematically proved the validity of calculating the biggest contribution from two projections and showed that expected behavior matches the empirical observations in the plots. Third, we have illustrated that the Feature Clock is applicable to the analysis of the classical non-linear dimensional reduction and the neural networks’ latent space.

## ACKNOWLEDGMENTS

We are grateful to Sapar Charyyev for the comments on the mathematical formulation of the problem. We thank Sebastian Baunsgaard and Florian Barkmann for their comments on the early drafts and for engaging in helpful discussions.

## REFERENCES

- [1] V. Berisha, C. Krantsevich, P. R. Hahn, S. Hahn, G. Dasarathy, P. Turaga, and J. Liss. Digital medicine and the curse of dimensionality. *npj Digital Medicine*, 4(1):153, 2021. doi: 10.1038/s41746-021-00521-5 1
- [2] A. Chatzimparmpas, R. M. Martins, and A. Kerren. t-viSNE: Interactive Assessment and Interpretation of t-SNE Projections. *IEEE Transactions on Visualization and Computer Graphics*, 26(8):2696–2714, 2020. doi: 10.1109/TVCG.2020.2986996 2
- [3] M.-Y. Cheng and H. tieng Wu. Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association*, 2013. doi: 10.1080/01621459.2013.827984 3
- [4] S. Cheng and K. Mueller. The data context map: Fusing data and attributes into a unified display. *IEEE Transactions on Visualization and Computer Graphics*, 22:121–130, 2016. doi: 10.1109/TVCG.2015.2467552 2
- [5] A. F. e. a. Connors. A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). *JAMA*, pp. 1591–1598, 1995. doi: 10.1001/jama.1995.03530200027032 3
- [6] K. Eckelt, A. Hinterreiter, P. Adelberger, C. Walchshofer, V. Dhanoa, C. Humer, M. Heckmann, C. Steinparz, and M. Streit. Visual exploration of relationships and structure in low-dimensional embeddings. *IEEE Transactions on Visualization and Computer Graphics*, 29(7):3312–3326, 2023. doi: 10.1109/TVCG.2022.3156760 2
- [7] R. Faust, D. Glickenstein, and C. Scheidegger. Dimreader: Axis lines that explain non-linear projections. *IEEE Transactions on Visualization and Computer Graphics*, pp. 481–490, 2019. doi: 10.1109/TVCG.2018.2865194 2
- [8] R. Fisher. Iris flower dataset, 1936. 1, 3
- [9] T. Fujiwara, O. Kwon, and K. Ma. Supporting analysis of dimensionality reduction results with contrastive learning. *IEEE Transactions on Visualization and Computer Graphics*, 26(01):45–55, 2020. doi: 10.1109/TVCG.2019.2934251 2
- [10] K. R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 3:453–467, 1971. doi: 10.2307/2334381 1
- [11] G. Gorin, M. Fang, T. Chari, and L. Pachter. Rna velocity unraveled. *bioRxiv*, 2022. doi: 10.1101/2022.02.12.480214 4
- [12] P. Indyk and R. Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *ACM Symposium on Theory of Computing*, p. 604–613, 1998. doi: 10.1145/276698.276876 1
- [13] Y. Jaradat, M. Masoud, I. Jannoud, et al. A Tutorial on Singular Value Decomposition with Applications on Image Compression and Dimensionality Reduction. In *International Conference on Information Technology*, 2021. doi: 10.1109/ICIT52682.2021.9491732 1
- [14] P. Joia, F. Petronetto, and L. G. Nonato. Uncovering Representative Groups in Multidimensional Projections. *Computer Graphics Forum*, 2015. doi: 10.1111/cgf.12640 2
- [15] K. Karuathna. Melody metrics: Decoding song popularity, 2023. doi: 10.34740/KAGGLE/DS/3914464 7
- [16] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, pp. 79–86, 1951. doi: 10.1214/aoms/1177729694 1
- [17] C. Li, M. Virgilio, K. Collins, and J. Welch. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. *Nature Biotechnology*, 41:1–12, 10 2022. doi: 10.1038/s41587-022-01476-y 1
- [18] Z. Li and Y. Qiu. Feature selection based on improved principal component analysis. In *Conference on Algorithms, Computing and Machine Learning*, p. 188–192, 2023. doi: 10.1145/3590003.3590036 1
- [19] G. C. Linderman and S. Steinerberger. Clustering with t-SNE, provably. *SIAM Journal on Mathematics of Data Science*, 1 2:313–332, 2017. doi: 10.1137/18M1216134 1
- [20] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1249–1268, 2017. doi: 10.1109/TVCG.2016.2640960 2
- [21] L. Manduchi, R. Marcinkevičs, M. C. Massi, T. Weikert, A. Sauter, V. Gotta, T. Müller, F. Vasella, M. C. Neidert, M. Pfister, B. Stieltjes, and J. E. Vogt. A deep variational approach to clustering survival data. In *International Conference on Learning Representations*, 2022. 1
- [22] W. E. Marcílio-Jr and D. M. Eler. Explaining dimensionality reduction results using shapley values. *Expert Systems with Applications*, 178:115020, 2021. doi: 10.1016/j.eswa.2021.115020 2
- [23] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, 2018. 1, 3
- [24] K. R. Moon, D. van Dijk, S. G. Zheng Wang, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37:1482–1492, 2019. doi: 10.1038/s41587-019-0336-3 1
- [25] C. Nefel et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell*, 178(4):835–849.e21, 2019. doi: 10.1016/j.cell.2019.06.024 6
- [26] M. Prummer, A. Bertolini, L. Bosshard, F. Barkmann, J. Yates, V. Boeva, T. T. P. Consortium, D. Stekhoven, and F. Singer. scROSHI: robust supervised hierarchical identification of single cells. *NAR Genomics and Bioinformatics*, 2023. doi: 10.1093/nargab/lqad058 1
- [27] T. Sainburg, L. McInnes, and T. Q. Gentner. Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning. *ArXiv e-prints*, 2021. 1
- [28] J. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes. Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus. *Annual Symposium on Computer Applications in Medical Care*, 10, 11 1988. 4, 7
- [29] J.-T. Sohns, M. Schmitt, F. Jirasek, H. Hasse, and H. Leitte. Attribute-based explanation of non-linear embeddings of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, pp. 540–550, 2022. doi: 10.1109/TVCG.2021.3114870 2
- [30] J. Stahneke, M. Dörk, B. Müller, and A. Thom. Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):629–638, 2016. doi: 10.1109/TVCG.2015.2467717 2
- [31] C. Turkey, A. Lundervold, A. J. Lundervold, and H. Hauser. Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2621–2630, 2012. doi: 10.1109/TVCG.2012.256 2
- [32] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 1, 3
- [33] M. Wattenberg, F. Viégas, and I. Johnson. How to use t-sne effectively. *Distill*, 2016. doi: 10.23915/distill.00002 1
- [34] J. Yates and V. Boeva. Deciphering the etiology and role in oncogenic transformation of the CpG island methylator phenotype: a pan-cancer analysis. *Briefings in Bioinformatics*, 2022. doi: 10.1093/bib/bbab610 1
- [35] J. Ye, R. Janardan, and Q. Li. Gpca: An efficient dimension reduction scheme for image compression and retrieval. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 354–363, 2004. doi: 10.1145/1014052.1014092 1
- [36] H. Zhou, Y. Zhang, Y. Zhang, and H. Liu. Feature selection based on conditional mutual information: minimum conditional relevance and minimum conditional redundancy. *Applied Intelligence*, 2019. doi: 10.1007/s10489-018-1305-0 1

## A SUPPLEMENTARY MATERIALS

In this document, we present mathematical proof of the Feature Clock<sup>2</sup> and further applications of Global, Local and Inter-group Clocks. Due to page constraints, we could not include them in the full size in the paper.

### A.1 Theoretical Foundation

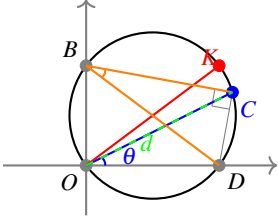


Figure 7: Single projection

We want to formally prove that clock technique is valid, and resulting perfect circles are the expected behavior (see Figure 8). For simplicity, it is enough to show that transformation (projection and rotation) results in circles for one high-dimensional feature.

First, the low-dimensional data  $X \in \mathbb{R}^{n \times 2}$  is projected on the radius of the semicircle at angles from  $0^\circ$  to  $180^\circ$  with  $5^\circ$  increments. For some projection angle  $\theta$ ,  $\cos(\theta) = \frac{d}{\|X\|} = \frac{X \cdot v}{\|X\| \|v\|}$  where  $v = (\cos \theta, \sin \theta)$  is a unit vector ( $\|v\| = 1$ ), and  $d$  is the distance vector from the origin of the semicircle to the projected points. It is obvious that  $d = X \cdot v$ ,  $d \in \mathbb{R}^{n \times 1}$ ,  $X \in \mathbb{R}^{n \times 2}$ ,  $v \in \mathbb{R}^{2 \times 1}$ .

Second, for each  $\theta$ , linear regression (LR) is used to predict projection  $d$  from the high-dimensional data  $S \in \mathbb{R}^{n \times d}$ , with a closed form solution  $\beta_\theta \in \mathbb{R}^{d \times 1}$ :

$$\beta_\theta = (S^T S)^{-1} S^T d = (S^T S)^{-1} S^T X v = A X \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad (1)$$

where  $A = (S^T S)^{-1} S^T$ .  $\beta_\theta$  is a vector with each feature's LR coefficient for a projection at angle  $\theta$ .

It is known that three non-collinear points make a unique circle. It can be shown by constructing perpendicular bisectors through each of the line segments formed between three points that do not lie on the same line. All perpendicular bisectors intersect at one point and are all equidistant from that intersection point. Therefore, we get a circle.

We always project on  $0^\circ$  and  $90^\circ$  from origin. Therefore, as shown in Fig. 7, we choose three non-collinear points: Origin  $O$ ,  $B$  on the y-axis, and  $D$  on the x-axis. We need to prove that fourth point  $C$  lies on the circumcircle of  $O, B, D$  by showing that  $\angle BCD = 90^\circ$ . If Eq. (2) holds, then  $\angle BCD = 90^\circ$ , and  $O, B, C, D$  form a circle.

$$BD^2 = BC^2 + CD^2 = BO^2 + OD^2 \quad (2)$$

Let's choose a  $i$ -th row of  $\beta_\theta$  and  $A$ , scalar  $\beta_\theta^i$  is a coefficient of  $i$ -th feature for angle  $\theta$  and  $a_i = ((S^T S)^{-1} S^T)_i$ . From Eq. (1),  $\beta_{0^\circ}^i = a_i X \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  and  $\beta_{90^\circ}^i = a_i X \begin{bmatrix} 0 \\ 1 \end{bmatrix}$  for  $\theta = 0^\circ$  and  $\theta = 90^\circ$  respectively.

We define four points  $C = (\beta_\theta^i \cos \theta, \beta_\theta^i \sin \theta)$ ,  $B = (0, \beta_{90^\circ}^i)$ ,  $D = (\beta_{0^\circ}^i, 0)$ , and  $O = (0^\circ, 0^\circ)$ , then  $OD = \beta_{0^\circ}^i$  and  $BO = \beta_{90^\circ}^i$ .

$$\begin{aligned} OD^2 + BO^2 &= a_i X \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} X^T a_i^T + a_i X \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} X^T a_i^T \\ &= a_i X \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} X^T a_i^T + a_i X \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} X^T a_i^T \\ &= a_i X \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} X^T a_i^T \end{aligned} \quad (3)$$

$$\begin{aligned} BC^2 &= (-\beta_\theta^i \cos \theta)^2 + (\beta_{90^\circ}^i - \beta_\theta^i \sin \theta)^2 \\ &= \left( a_i X \begin{bmatrix} \cos \theta^2 \\ \sin \theta \cos \theta \end{bmatrix} \right)^2 + \left( a_i X \begin{bmatrix} 0 \\ 1 \end{bmatrix} - a_i X \begin{bmatrix} \cos \theta \sin \theta \\ \sin \theta^2 \end{bmatrix} \right)^2 \\ &= \left( a_i X \begin{bmatrix} \cos \theta^2 \\ \sin \theta \cos \theta \end{bmatrix} \right)^2 + \left( a_i X \begin{bmatrix} -\cos \theta \sin \theta \\ \cos \theta^2 \end{bmatrix} \right)^2 \\ &= \cos^2 \theta \left( a_i X \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \right)^2 + \cos^2 \theta \left( a_i X \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} \right)^2 \end{aligned} \quad (4)$$

$$\begin{aligned} CD^2 &= (\beta_{0^\circ}^i - \beta_\theta^i \cos \theta)^2 + (-\beta_\theta^i \sin \theta)^2 \\ &= \left( a_i X \begin{bmatrix} 1 \\ 0 \end{bmatrix} - a_i X \begin{bmatrix} \cos \theta^2 \\ \sin \theta \cos \theta \end{bmatrix} \right)^2 + \left( a_i X \begin{bmatrix} \sin \theta \cos \theta \\ \sin \theta^2 \end{bmatrix} \right)^2 \\ &= \left( a_i X \begin{bmatrix} \sin \theta^2 \\ -\sin \theta \cos \theta \end{bmatrix} \right)^2 + \left( a_i X \begin{bmatrix} \sin \theta \cos \theta \\ \sin \theta^2 \end{bmatrix} \right)^2 \\ &= \sin^2 \theta \left( a_i X \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix} \right)^2 + \sin^2 \theta \left( a_i X \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \right)^2 \end{aligned} \quad (5)$$

We combine Eq. (4) and Eq. (5), and  $\sin^2 \theta + \cos^2 \theta = 1$ :

$$\begin{aligned} BC^2 + CD^2 &= \left( a_i X \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \right)^2 + \left( a_i X \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix} \right)^2 \\ &= a_i X \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} X^T a_i^T \\ &\quad + a_i X \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix} \begin{bmatrix} \sin \theta & -\cos \theta \end{bmatrix} X^T a_i^T \\ &= a_i X \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix} X^T a_i^T \\ &\quad + a_i X \begin{bmatrix} \sin^2 \theta & -\sin \theta \cos \theta \\ -\sin \theta \cos \theta & \cos^2 \theta \end{bmatrix} X^T a_i^T \\ &= a_i X \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} X^T a_i^T \end{aligned} \quad (6)$$

We can see that Eq. (3) and Eq. (6) are equal,  $\angle BCD = 90^\circ$ , point  $C$  lies on the circumcircle, and, therefore, a circle is formed by projecting and rotating.

We know that  $\angle BOD = 90^\circ$  and  $BD$  is the diameter of a circle. Since the biggest contribution  $OK$  defines the circle, it is also the diameter of the circle. The intersection of  $OK$  and  $BD$  is the center of a circle. Therefore, we can find coordinates of  $K$  after projecting and rotating on the x- and y-axis only, at angles of  $0^\circ$  and  $90^\circ$  respectively. If  $D = (x, 0)$  and  $B = (0, y)$ , then  $K = (x, y)$ .

### A.2 Visualization of all coefficients

If we project on lines with  $\theta$  iterates, we get perfect circles as shown in Fig. 8.

### A.3 Nefitel Dataset

We use a dataset popularly called Nefitel [25]. The dataset includes a single-cell analysis of glioblastoma, a type of cancer that starts as a growth of cells in the brain or spinal cord. The dataset provides a blueprint for glioblastoma, integrating the malignant cell programs and their modulation by genetic drivers. There are six glioblastoma state signatures for malignant cells included in the data: NPCLike1, NPCLike2, MESlike1, MESlike2, ACLike, and OPLike. Given all signatures and the number of expressed genes, biologists try to detect which cells (data points) are characterized by which signature.

<sup>2</sup><https://pypi.org/project/feature-clock>

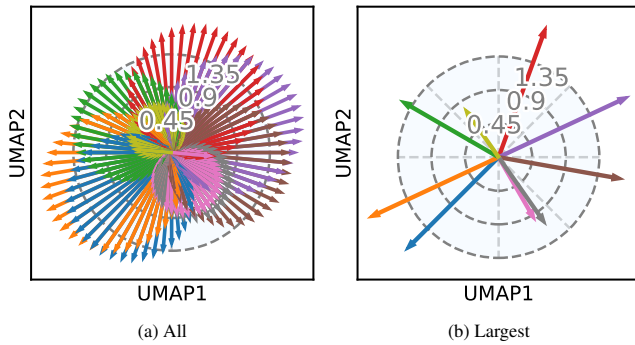


Figure 8: All vs Largest Coefficients

A usual approach to compare signatures is to plot low-dimensional data for each signature color-coding data point with their high-dimensional values, as shown in the rightmost plot in Fig. 9. Another method is to assign labels to each point and to visualize low-dimensional data color-coded by these labels. Labels are created by assigning the name of a signature with maximum value. With this method, a user can see if there are any clusters or structures in the data.

Fig. 9 shows global, local, and inter-group clocks. The Global Clock gives a compact overview of general trends in the data. We see that the clock in Appendix A.3a shows how values change and that, e.g., OPCLike signatures have the biggest impact in approximately 70° direction. The same trend can be seen in scatter plots.

Two local clocks analyze embeddings at a finer granularity. We use HDBSCAN to find clusters in high-dimensional data. There are two clusters, and, therefore, two clocks. For the big cluster on the left, the direction of impact of the OPCLike signatures slightly changes compared to the big clock. In the second clock (on the right), an OPCLike arrow points in the direction of 150°. Properties of points of the right cluster differ from the left one, this leads to completely different directions and higher coefficients (arrow length marked on circle). This can be interpreted in connection with the Global Clock where the effect of combining impacts of both clusters is visualized. Although the right cluster has bigger coefficients, it is impact on the Global Clock is leveled out because of the smaller amount of points. Such details as the impact within a single cluster can not be seen on a bigger scale and even when plotting all features.

Fig. 9c, the Inter-group Clock, shows the relationship between two HDBSCAN clusters and which features can be changed to move from the center of one cluster to the center of another. It is visible that the most significant feature between the two clusters is NPCLike2. NPCLike2 arrow points in the direction of a smaller cluster (on the right). This means that to move from a bigger cluster to a smaller one, we need to increase NPCLike2 impact. The same trend can be seen in the rightmost plot in the Fig. 9. Additionally, scatter plots justify that the biggest value difference between the points is in NPCLike2 signatures.

#### A.4 Pima Indians Diabetes Database

The Pima Indians Diabetes Database [28] originates from the National Institute of Diabetes and Digestive and Kidney Diseases. The goal of the dataset is to predict if a patient has diabetes based on personal data, e.g., blood pressure and skin sickness. We use actual class labels for the Local and Inter-group Clocks.

In the Global Clock, Fig. 10a, we see that blood pressure, age, glucose, and insulin level influence diabetes among Pima females. Interestingly, skin thickness varies among all observations, but body mass index (BMI) is insignificant.

Fig. 10b shows Local Clocks for two classes. Both clocks are driven by almost the same features but the directions differ. Addi-

tionally, BMI is insignificant for the both classes.

Inter-group Clock highlights that when pedigree, BMI, and glucose increase, the likelihood of having diabetes is higher.

#### A.5 Melody Metrics: Decoding Song Popularity

A synthetic dataset [15] from Kaggle is designed to unravel the mystique behind song popularity. The task is to predict song popularity (regression) based on the information about it. We create two classes from of the original target *Popularity* (see Fig. 11d), and visualize a clocks for all songs.

From the Global Clock in Fig. 11a, it is clear that lyrical content and a more recent year of release distinguish popular songs. Number of instruments has the highest variability among all observations. Genre, song length, and tempo influence music towards lower popularity, but contributions of these features are significantly lower compared with a number of instruments.

Local Clocks, Fig. 11b, show that same features drive both classes but some variable point in different directions.

The Inter-group Clock in Fig. 11c provides insights into what distinguishes unpopular melodies from hits. We can "move" between two classes by changing lyrical content, tempo, year of release, or number of instruments. Additionally, it appears that changes of genre are insignificant between two classes.

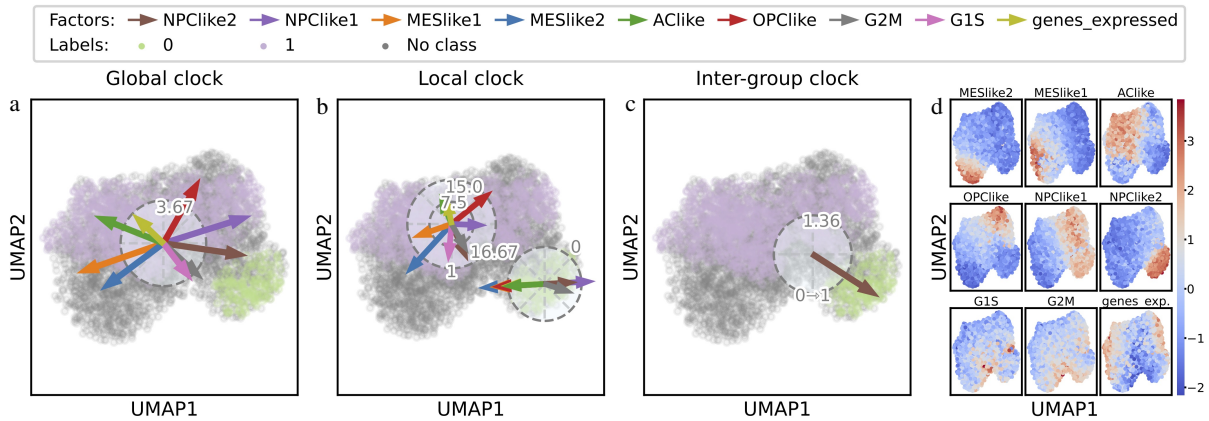


Figure 9: (a) The Global Feature Clock shows all significant glioblastoma signatures for all observations. (b) Local Clocks analyze the data at the granularity of HDBSCAN clusters, and (c) Inter-group Clock explains how features change between two clusters. (d) Scatter plots help to validate Feature Clocks and visualize a plot per each high-dimensional feature.

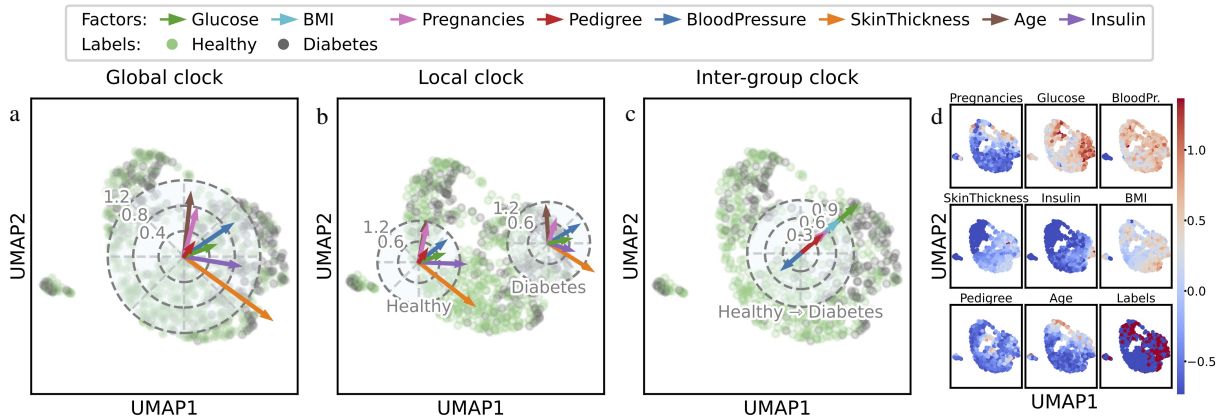


Figure 10: (a) The Global Feature Clock highlights features that differentiate people with and without diabetes, while (b) Local Clocks show how features range within original classes. (c) Inter-group Clock describes how personal data differs between females with and without diabetes. (d) Scatter plots are intended to support clocks and ease the users' perception of a new visualization.

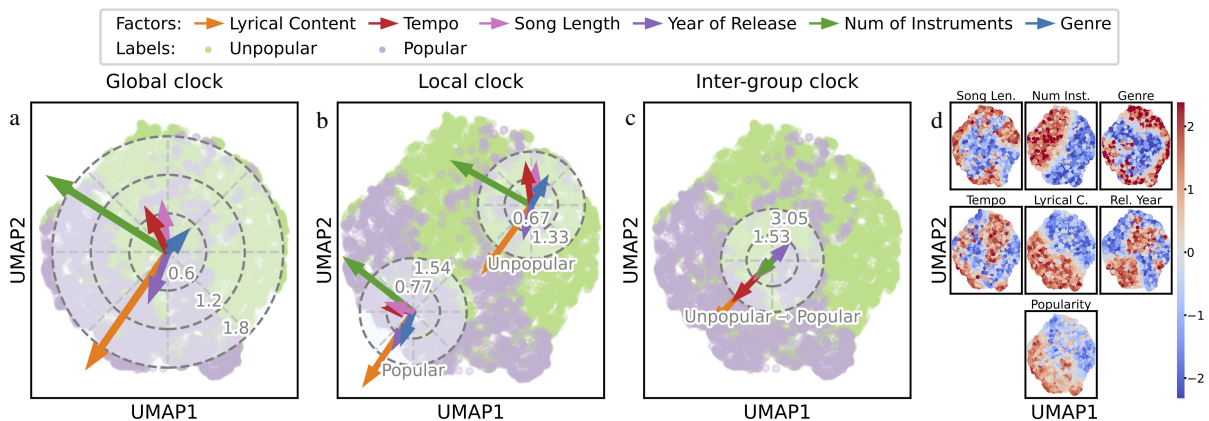


Figure 11: To explore melody popularity, we visualize (a) a global clock for all songs, (b) a Local Clock for each class, and (c) an inter-group clock between two classes. (d) For each feature, we show a scatter to confirm the correct behaviour of clocks where *Popularity* is the original continuous feature. We transformed *Popularity* into two classes for the clocks.