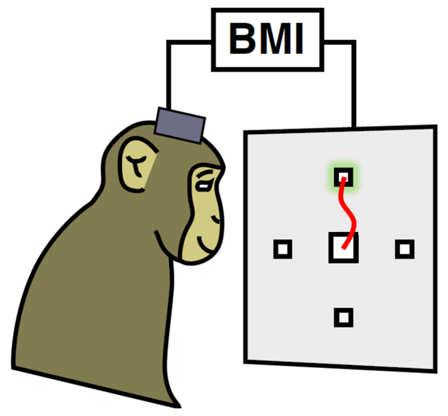



# Reducing the Dimensionality of a Monkey Neuron Dataset

Benjamin Pellow
Oral: https://www.imperial.ac.uk/mathematics

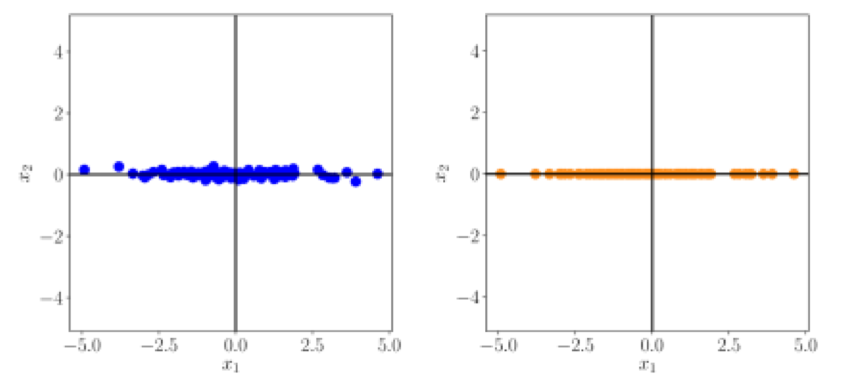
Imperial College London

## Motivation for PCA

Principal component analysis (PCA) is a technique used to reduce the dimensionality of data, to make analysis easier, whilst retaining as much of the variation of the original data as possible.

Performing PCA involves transforming the data to a new set of variables known as principal components (PCs).

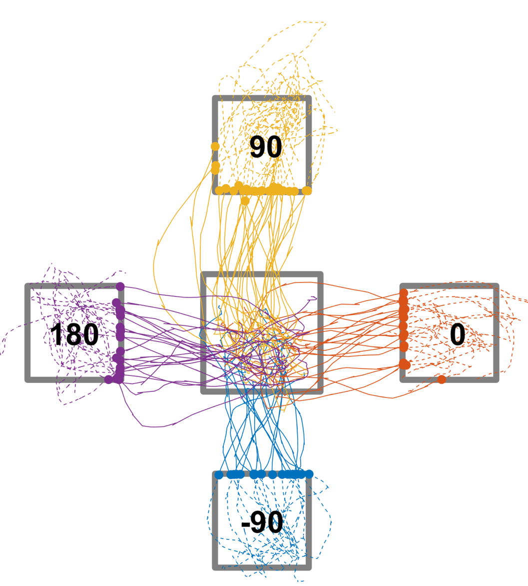
## Maths of PCA



(a) Dataset with $x_1$ and $x_2$ coordinates. (b) Compressed dataset where only the $x_1$ coordinate is relevant. [1]

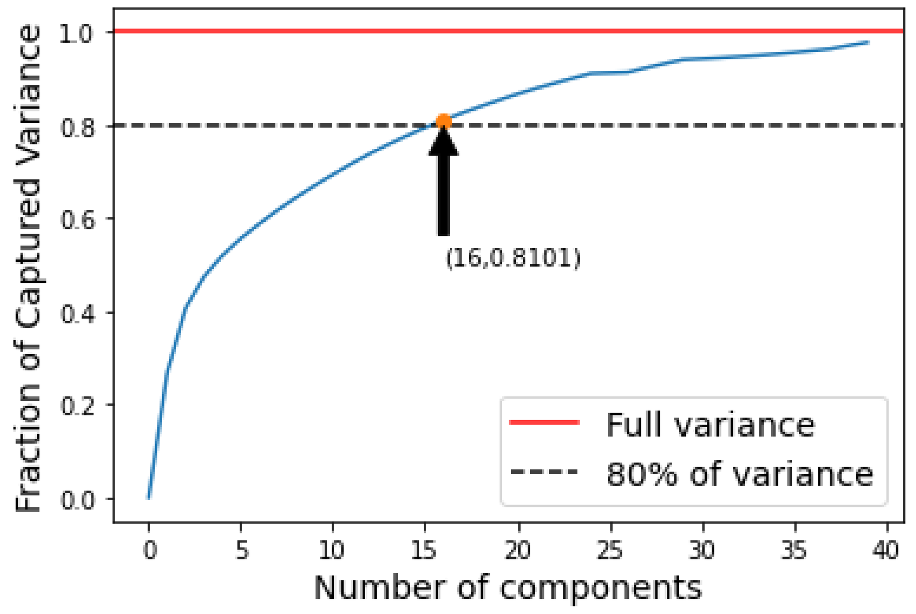
1. Given $N$, $p$-dimensional vectors $\mathbf{x}^{(n)}$ which are the data points (as an $N \times p$ matrix)
2. First, we need the vector $\phi_1^*$ which maximises the variance of data projections along it
3. $\phi_1^*$ turns out to be the eigenvector of $C_{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}^{(n)} - \boldsymbol{\mu})(\mathbf{x}^{(n)} - \boldsymbol{\mu})^T$ with the largest eigenvalue.
4. Hence to perform PCA, we find all the eigenvectors of $C_{\mathbf{x}}$ and order them from largest to smallest based on their respective eigenvalues (all of which are $\geq 0$ since $C_{\mathbf{x}}$ is positive semi-definite).
5. The lower dimensional representation of the data is the projection onto the first $k$ of these eigenvectors with $k << p$

## The dataset




**88 trials** are recorded, in which a monkey moves a cursor to one of four targets (at 0, 90, 180, and -90 degrees) using a brain machine interface (BMI). In each trial the activity of **43 neurons** is recorded over **200 time steps**. [2]

## How many principal components?



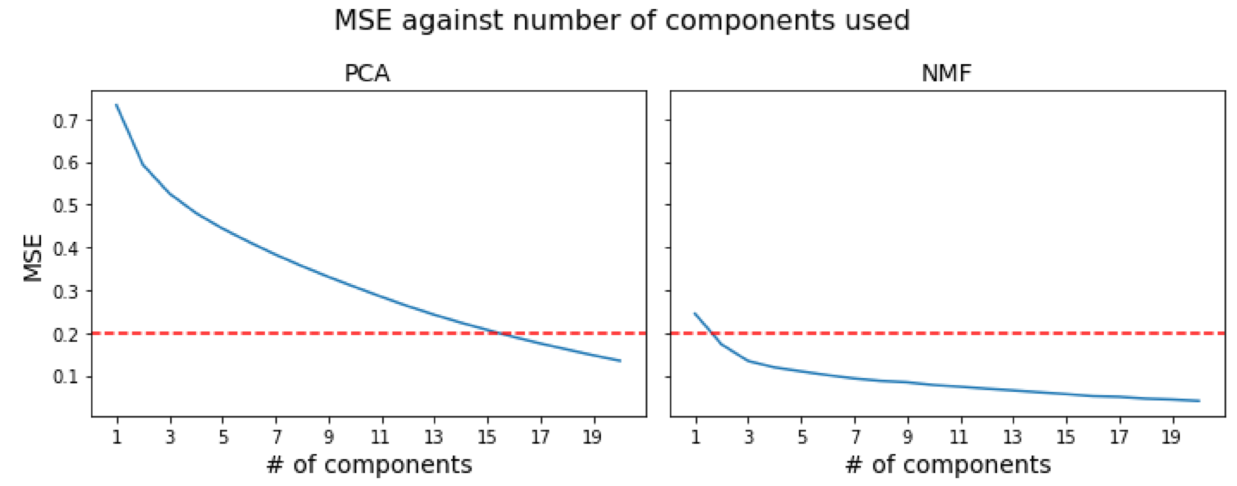

To decide the number of components to use we set a threshold for the captured variance then make the decision based on this.

Above is a plot of captured variance against number of components used, for the the monkey BMI dataset at time step $t = 50$ . If we require at least 80% of the variance to be captured then we need to use 16 components.

## Comparing PCA with NMF

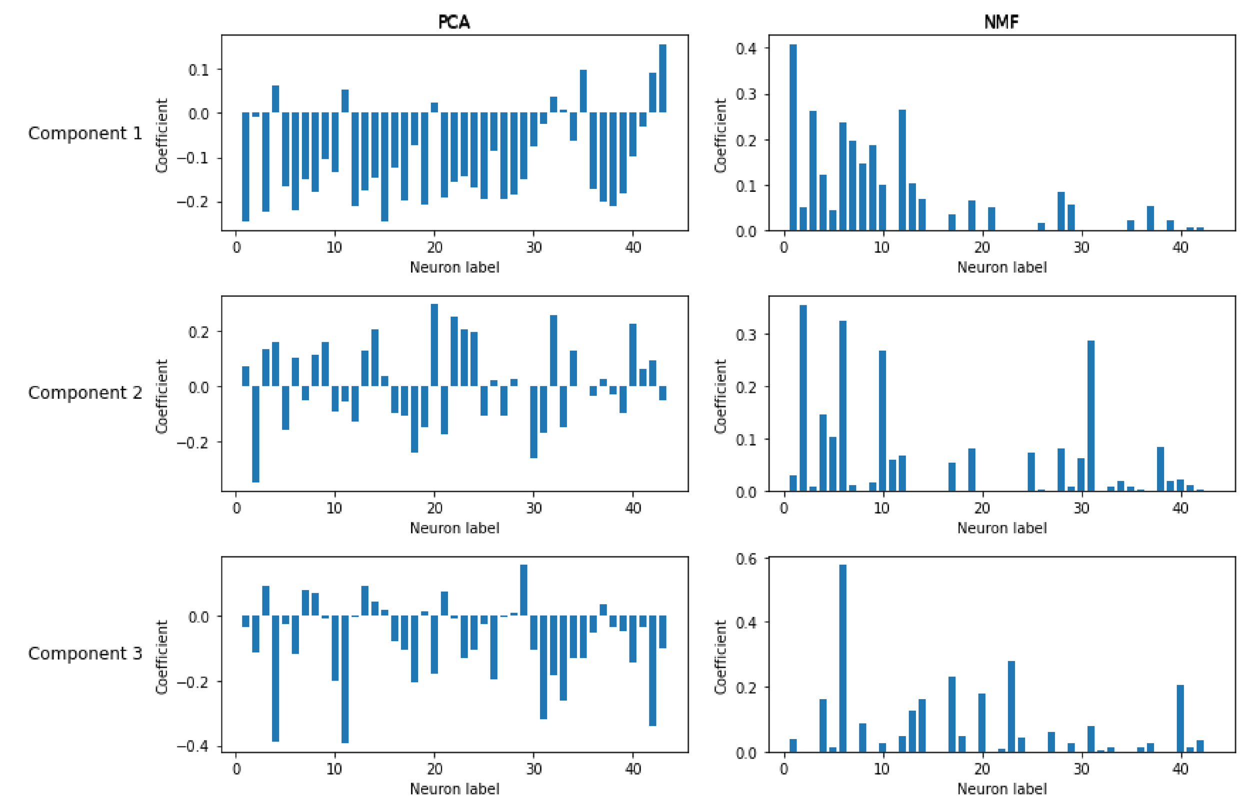
*Non-Negative Matrix Factorisation (NMF)* is another dimensionality reduction technique, unlike PCA it requires the original data to be non-negative.

NMF involves factorising the original data matrix $X_{(N \times p)}$ into a product of two matrices $W_{(N \times r)}$ and $H_{(r \times p)}$ where $r << p$ and whose elements are all non-negative. $W$ and $H$ can be computed using Lee and Seung's multiplicative update rule [4]




The red line shows a Mean-Squared Error (MSE) of 20%. Using NMF with these data requires fewer components than PCA to achieve an error of 20%.

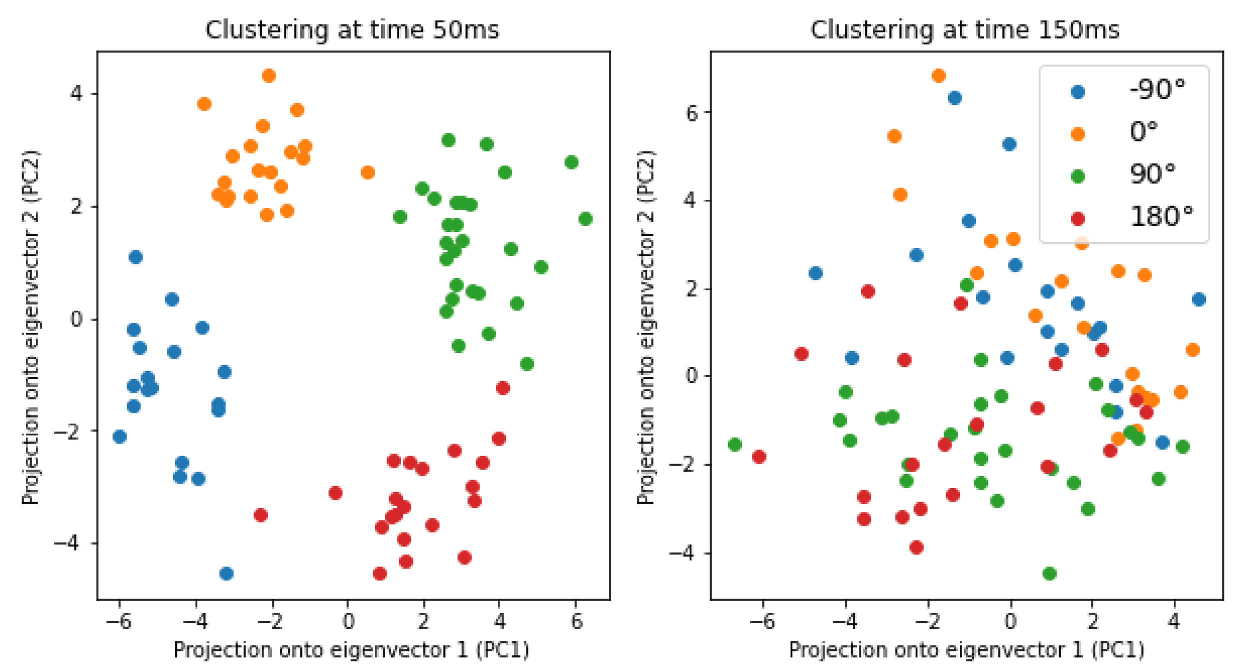
## Interpreting the components




Here I have performed PCA and NMF with 3 components at $t = 50$ ms and plotted the values of each for the 43 neurons in order to visualise each component. They can be seen as 'modes' - weighted groupings of neurons - involved with the movements at this time step.

Since the NMF components are constrained to be positive, interpreting them is easier. They also have sparser representations compared to the PCs, making the associated modes of neural activity clearer to see.

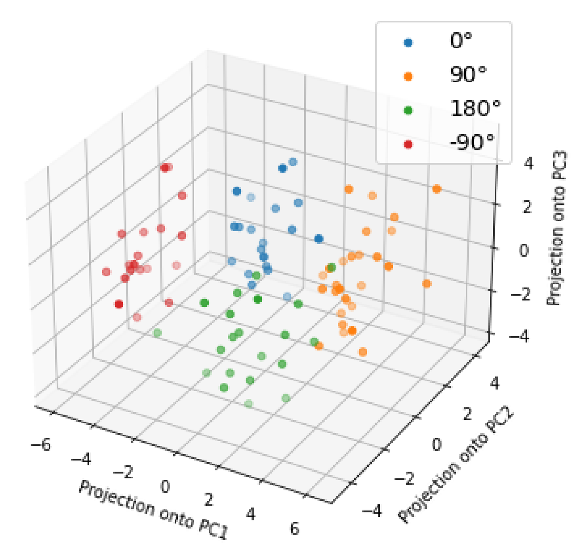
## Clustering of the data at 2 time points




At times around $t = 50$ms, where the monkey is moving the cursor, the plot exhibits clear clusters of same-angle trials. This shows that there are particular 'modes' of neuronal activity associated with the movements to particular cursor angles.

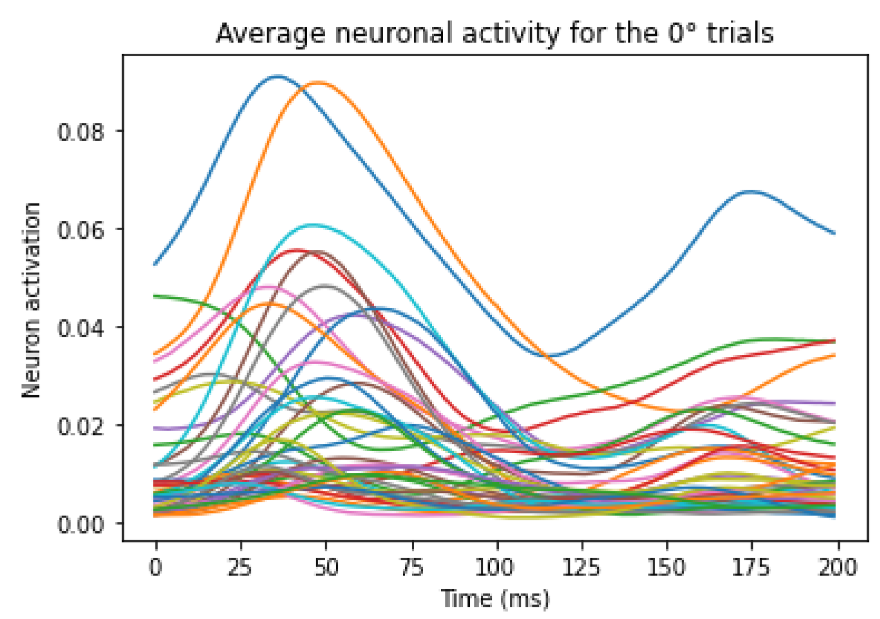
When the cursor is being held stationary, for example for $t > 100$ms when the monkey has reached its target, the clustering fades. Holding the cursor stationary is less unique to a particular angle, hence the 'modes' of neuronal activity are less distinct.

## More on clustering

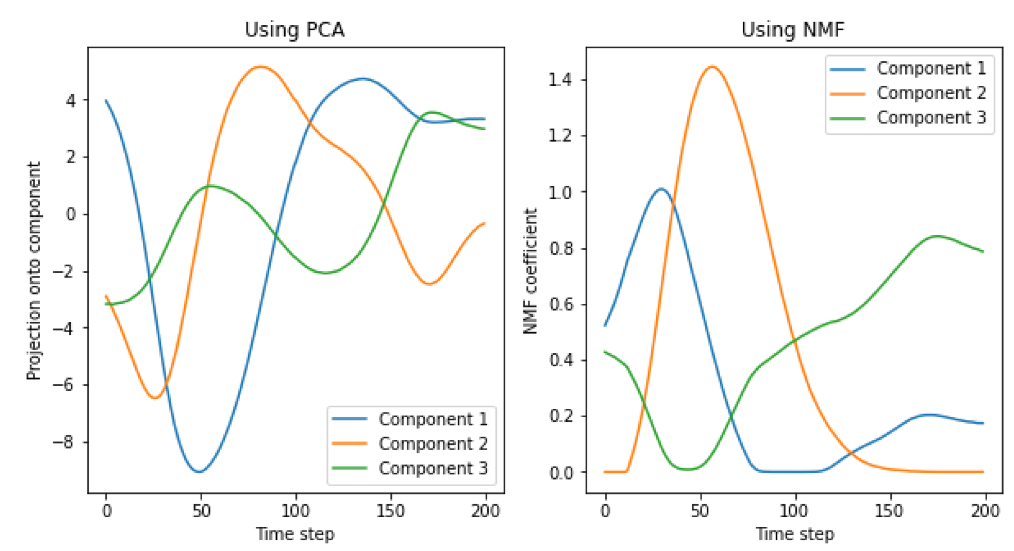



By plotting the projections of the data onto 3 PCs, we can visualise the clustering behaviour in 3D. Interestingly, using NMF around $t = 50$ms and plotting the rows of $W$ also shows clustering, albeit less pronounced than with PCA.

## Averaging over the 0° trials




$\Rightarrow$




Using PCA and NMF on the average of the 0° trials, the right plots show the projections of the data onto each component. Note that the NMF plot shows a sparser representation with sharper peaks.

## References

[1] M.P. Deisenroth, A.A. Faisal, and C. Soon Ong. *Mathematics for Machine Learning - chapter 10.* Cambridge University Press; 2020.

[2] T.G. Kolda *Monkey BMI Tensor Dataset* https://gitlab.com/tensors/tensor_data_monkey_bmi [Accessed: 04/06/2022]

[3] D.D. Lee and H.S. Seung. *Learning the parts of objects by non-negative matrix factorization.* Nature. 1999; 401

[4] D.D. Lee and H.S. Seung. *Algorithms for Non-negative Matrix Factorization* 2-3. https://papers.nips.cc/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf