The University Of Queensland AUSTRALIA | CREATE CHANGE

Pattern Analysis

Principal Component Analysis

Dr Shekhar “Shakes” Chandra

V1.0

> “The methods of theoretical physics should be applicable to all those branches of thought in which the essential features are expressible with numbers.”
>
> Paul M. Dirac (link)
> (1902-1984)

Change of Basis

- Data/Signals as a Vector with some basis $\mathbf{e}_i$

$$x(t) = \sum_{i}^{N} a_i \mathbf{e}_i$$

- We can change the basis to a more convenient set $\boldsymbol{\alpha}_i$ by projecting the data to these new basis set

$$\hat{x}(t) = \sum_{i}^{N} a_i (\mathbf{e}_i \cdot \boldsymbol{\alpha}_i)$$

- These bases can be predefined, such as trigonometric functions resulting in the Fourier Transform. Or we can compute new ones based on some property of the data/signal, such as the variance with respect to the mean, resulting in Eigen-decomposition.

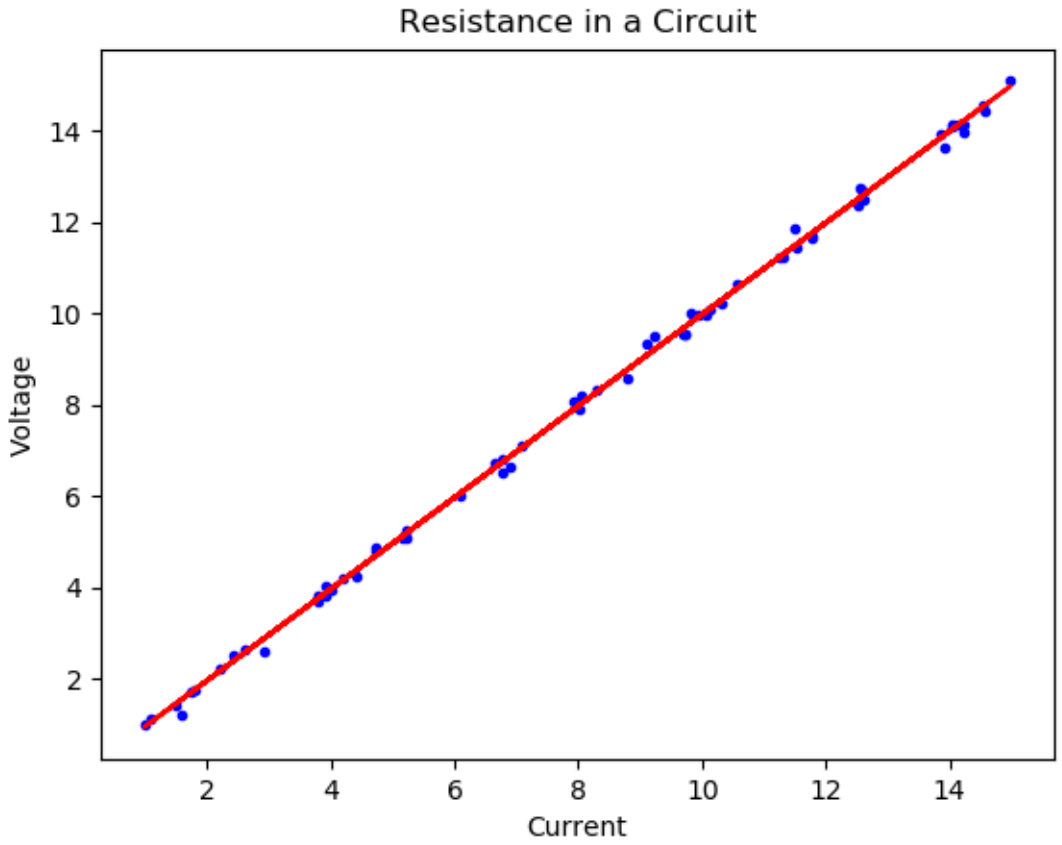
Example

- Set of points is measured and found to have the relationship on the right
- Each point has a pair of values that make a vector $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$
- If we place the vectors into matrix stacked vertically with M columns

$$\boldsymbol{X} = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 & \cdots & x_{N-1} \\ y_0 & y_1 & y_2 & y_3 & & y_{N-1} \end{bmatrix}$$

- Ideally, we want the most "compact" representation so that the resulting number of columns $k \ll M$.
- This way, we would like a new set of vectors $\boldsymbol{u}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$, so that $0 \leq i < k$ will be

$$\boldsymbol{U} = \begin{bmatrix} u_0 & u_1 & u_2 & u_3 & \cdots & u_{k-1} \\ v_0 & v_1 & v_2 & v_3 & & v_{k-1} \end{bmatrix}$$



Example

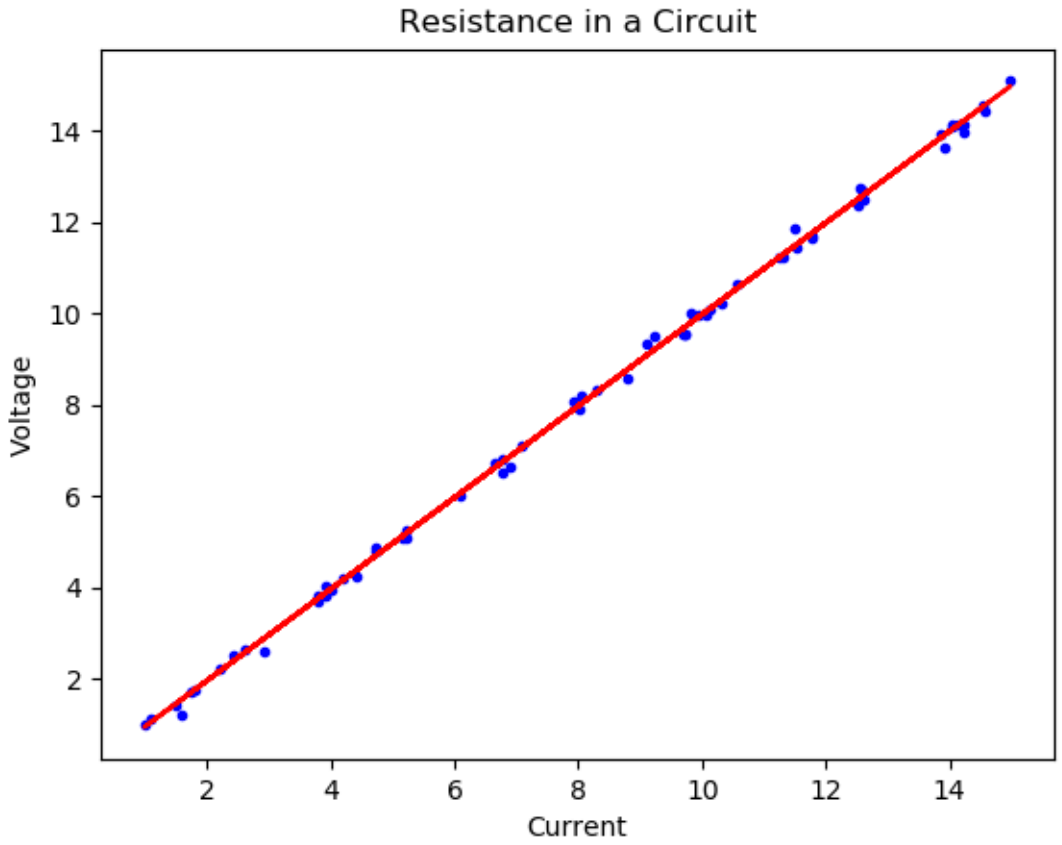
- We have a set of 64 points measured and found to have the relationship on the right
- We can construct a data matrix $\boldsymbol{X}$ with vectors $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ stacked vertically with M columns

$$\boldsymbol{X} = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 & \cdots & x_{N-1} \\ y_0 & y_1 & y_2 & y_3 & & y_{N-1} \end{bmatrix}$$

- If we ignore the small "noise" in our data points, our most "compact" representation is that of a line, so that the resulting number of columns $k \ll M$, i.e. $k = 1$.
- Our new set of vectors $\boldsymbol{u}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$ will be

$$\boldsymbol{U} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}$$

- To get any one of the points in our original $\boldsymbol{X}$, we simply need to scale our vector $\boldsymbol{u}$ by a scale factor b per point, so that the reconstructed point $\tilde{x}_i$ is $b_i \boldsymbol{u}$



Another Example

- We have three sets of 128, 64 and 32 points respectively measured and found to have the relationship on the right
- We can construct a data matrix $\boldsymbol{X}$ with vectors $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ stacked vertically with M columns of all three sets where

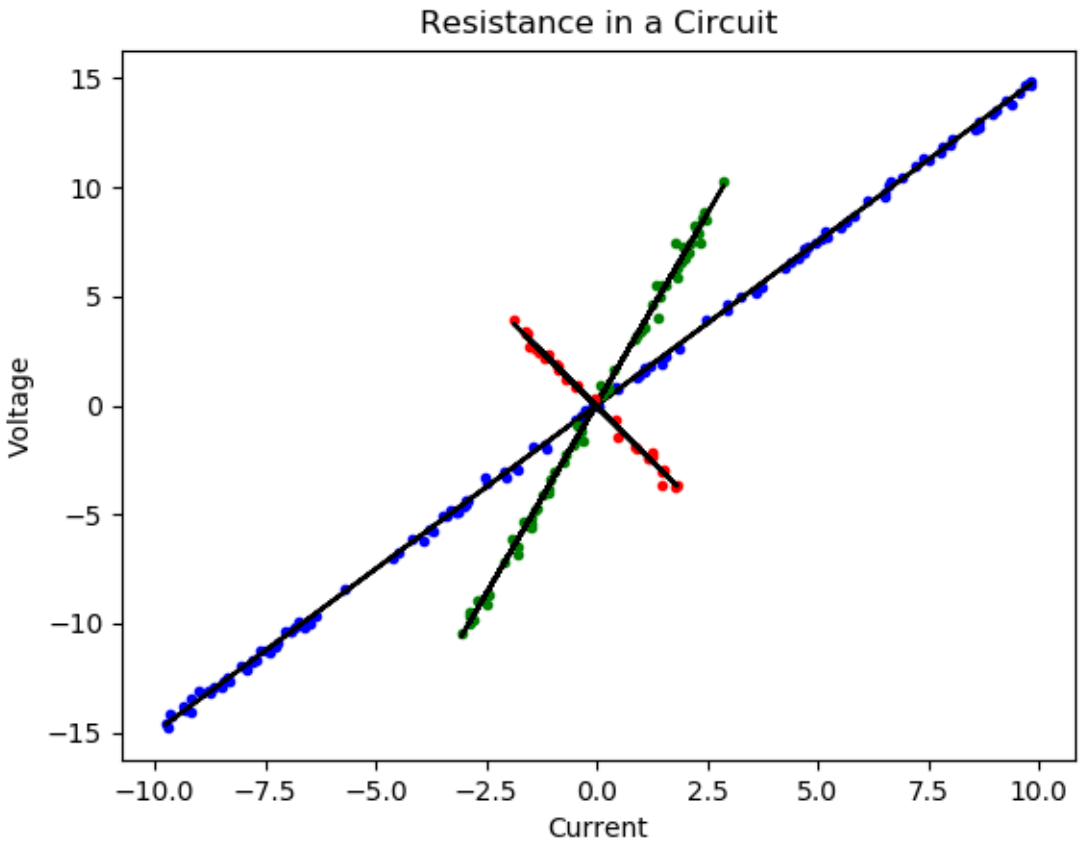
$$\boldsymbol{X} = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 & \cdots & x_{N-1} \\ y_0 & y_1 & y_2 & y_3 & & y_{N-1} \end{bmatrix}$$

- If we ignore “noise” in our data points, our most “compact” representation is that of three lines, so that the resulting number of columns $k \ll M$, i.e. $k = 3$.
- Our new set of vectors $\boldsymbol{u}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$ will be

$$\boldsymbol{U} = \begin{bmatrix} u_0 & u_1 & u_2 \\ v_0 & v_1 & v_2 \end{bmatrix}$$

- To get any one of the points in our original $\boldsymbol{X}$, we simply need to construct a linear combination of our vectors $\boldsymbol{u}$ scaled by a scale factor b per point, so that the reconstructed point $\tilde{x}_i$ is

$$\tilde{x}_i = \sum_{i=0}^{k-1} b_i \mathbf{u}_i$$



Eigen-decomposition

- In general, we need to do this for arbitrary data sets, where these simple relationships are not necessary apparent.
- Ideally, our data would be represented using the basis set that gives us the most “compact” representation.
- By compact we mean representing our data with as few basis vectors as possible.
- This compact representation can only be obtained by identifying the patterns in our data, its principal components.
- This representation involving principal components is called principal component analysis (PCA).
- A form of PCA was first introduced by Pearson [1901] and also known as the Karhunen–Loève Transform.
- Recall that in the chapter on symmetry, we defined symmetry as a property that remains unchanged after a transformation.
- Eigen-decomposition is the basis of PCA and provides a method to do this with respect to basis vectors of a given dataset

Eigen-decomposition

- In general, we need to do this for arbitrary data sets, where these simple relationships are not necessary apparent.
- Formally, we are trying to compute the eigen-decomposition

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}$$

- Here $\boldsymbol{A}$ is a transformation matrix, $\boldsymbol{u}$ is a vector called an eigenvector and λ is a scalar value called the eigenvalue.
- If we solve for the eigenvector $\boldsymbol{u}$, we get the expression

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{u} = 0$$

- Here $\boldsymbol{I}$ is the identity matrix and we solve for the non-trivial solutions of the equation.
- In other words, these eigenvectors $\boldsymbol{u}$ are vectors that remain invariant to any rotations induced by transformation $\boldsymbol{A}$. The vectors $\boldsymbol{u}$ do not change direction when $\boldsymbol{A}$ is applied to them.

Eigen-decomposition

- To compute this eigen-decomposition, we can use the singular value decomposition (SVD) algorithm that decomposes the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^{\boldsymbol{T}}$$

- Where $\boldsymbol{W}$ is a diagonal matrix whose entries are the singular values in descending order that correspond to the eigenvalues of the decomposition
- The matrix $\boldsymbol{U}$ corresponds to the eigenvectors of the decomposition


Eigenvalues

Conclusion

- The PCA finds the most compact representation of our data
- This compactness is often referred to as dimensionality reduction as it reduces our columns of our data matrix into a smaller number of vectors
- PCA can be used to model data into a lower dimensional vector space and is also sometimes referred to as sub-space learning
- PCA is often used as a “feature(s)” or to preprocess the data as input to more sophisticated algorithms for classification, such as random forests

What's Next?

How can we compute more complicated tasks such as classification and segmentation using the transforms and features we've learnt so far?

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Thank you

Dr Shekhar “Shakes” Chandra | Lecturer
School of Information Technology and Electrical Engineering
shekhar.chandra@uq.edu.au
+61 7 3365 8359

@shakes76
shekhar-s-chandra