Machine Learning in Geophysics

Lecture 5 – Principal Component Analysis

PCA

Principal Component Analysis is a data analysis technique used in many fields and known under a variety of names

- Karhunen-Loève transform (signal analysis)
- Empirical orthogonal functions (meteorology)
- Empirical component analysis

Closely related to Singular Value Decomposition (SVD), sometimes considered unsupervised ML technique.

SVD – Definition

Given a complex $M \times N$ matrix $\mathbf{M}$, we can find matrices $\boldsymbol{U}$, Σ, $\boldsymbol{V}$ such that

$$\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{V}^*$$

Where

- $\boldsymbol{U}$ is a $M \times M$ unitary matrix
- $\boldsymbol{V}$ is a $N \times N$ unitary matrix
- Σ is a $M \times N$ rectangular diagonal matrix with non-negative real values

Question

What is the defining property of a unitary matrix

SVD – Definition

Given a complex $M \times N$ matrix $\mathbf{M}$, we can find matrices $\boldsymbol{U}$, Σ, $\boldsymbol{V}$ such that

$$\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{V}^*$$

Where

- $\boldsymbol{U}$ is a $M \times M$ unitary matrix
- $\boldsymbol{V}$ is a $N \times N$ unitary matrix
- Σ is a $M \times N$ rectangular diagonal matrix with non-negative real values

Question

What is the defining property of a unitary matrix

Answer

$\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{U}^*\boldsymbol{U} = I$

SVD – Nomenclature

The entries of

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ 0 & 0 & \dots & \sigma_{pp} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

are the singular values $\sigma_{11}, \dots, \sigma_{pp}$ where $p = \min(m, n)$. The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left singular vectors and right singular vectors, respectively.

SVD and eigenvalues

$$\boldsymbol{M}\boldsymbol{M}^* = \boldsymbol{U}\Sigma\boldsymbol{V}^*(\boldsymbol{U}\Sigma\boldsymbol{V}^*)^* = \boldsymbol{U}\Sigma\boldsymbol{V}^*\boldsymbol{V}\Sigma^*\boldsymbol{U}^* = \boldsymbol{U}\Sigma\Sigma^*\boldsymbol{U}^*$$

and similarly

$$\boldsymbol{M}^*\boldsymbol{M} = \boldsymbol{V}\Sigma^*\Sigma\boldsymbol{V}^*.$$

This means $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{M}^*\boldsymbol{M}$ and $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{M}\boldsymbol{M}^*$

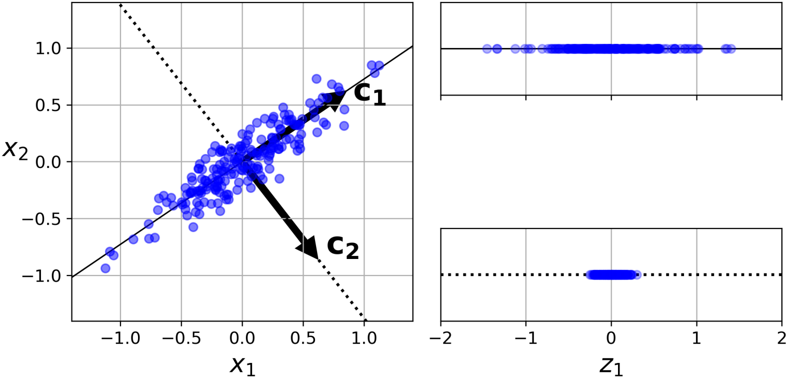
From SVD to PCA

- Have a set of features $\boldsymbol{x}_i$, $i = 1, \ldots, N$ with M values each and zero mean
- Construct matrix $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_N \end{pmatrix}$
- Calculate SVD of $\boldsymbol{M}$, columns of $\boldsymbol{V}$ are Principal Components
- Alternativ view: Calculate empirical covariance matrix

$$\boldsymbol{C} = \boldsymbol{M}^* \boldsymbol{M}$$

calculate eigenvalues of $\boldsymbol{C}$, eigenvectors are Principal components.

Simple example



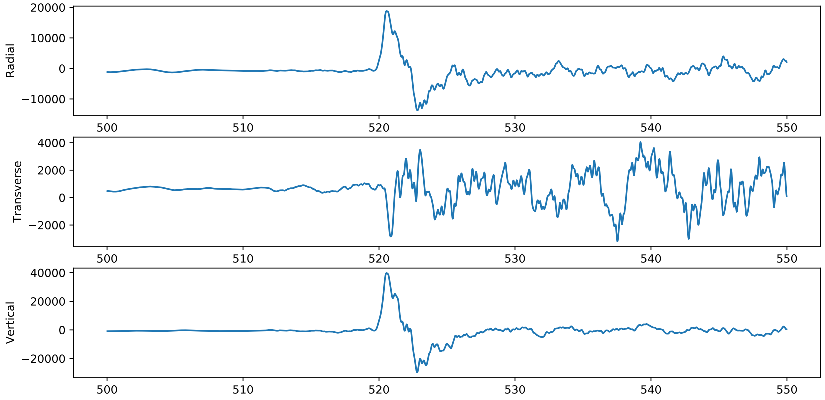
$$\Sigma = \begin{pmatrix} 6.96 & 0 \\ 0 & 1.53 \end{pmatrix} \qquad V = \begin{pmatrix} -0.79 & -0.6 \\ -0.6 & 0.79 \end{pmatrix}$$

Simple example

- Entries of Σ show us relative importance of Principal Components
- Columns of $\boldsymbol{V}$ show (orthogonal) directions
- Can project to new coordinates $\boldsymbol{X}_{proj} = \boldsymbol{XV}$

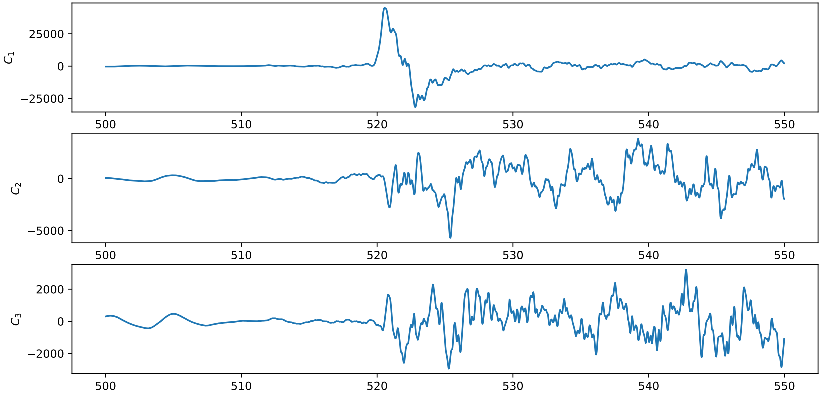
$$\Sigma = \begin{pmatrix} 6.96 & 0 \\ 0 & 1.53 \end{pmatrix} \qquad \boldsymbol{V} = \begin{pmatrix} -0.79 & -0.6 \\ -0.6 & 0.79 \end{pmatrix}$$

Feature extraction



Three component seismogram around direct P-wave arrival.

Feature extraction



Projected components. What has changed?

Explanation

$$V = \begin{pmatrix} 0.43 & -0.66 & -0.62 \\ -0.02 & 0.68 & -0.74 \\ 0.90 & 0.33 & 0.28 \end{pmatrix}$$

- $\boldsymbol{X}_{proj} = \boldsymbol{XV}$ means projected components are linear combinations of original data
- $\boldsymbol{V}$ unitary (orthogonal for real valued data) corresponds to rotations and reflections
- Transverse component does not contribute

Explanation

With only radial and vertical component

$$\boldsymbol{V} = \begin{pmatrix} 0.43 & -0.90 \\ 0.90 & 0.43 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \Rightarrow \theta = 64^\circ$$

Calculation of the incidence angle with the IASP91 model gives $\theta = 68^\circ$.

Dimensionality reduction

In our moment tensor classification problem

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}$$

we already exploited symmetry $M_{12} = M_{21}, \ldots$. This left us with 6 input quantities $M_{11}, M_{22}, M_{33}, M_{12}, M_{13}, M_{23}$. Can we reduce this further?

PCA

$$\text{diag}\Sigma = 131, 118, 101, 96, 94, 2.65$$

sometimes more intuitive to display in terms of explained variance ratio

$$EVR_i = \frac{\Sigma_{ii}^2}{\sum_i \Sigma_{ii}^2} = 0.29, 0.23, 0.17, 0.15, 0.15, 0.001$$

Question

What does this tell us?

PCA

$$\text{diag}\Sigma = 131, 118, 101, 96, 94, 2.65$$

sometimes more intuitive to display in terms of explained variance ratio

$$EVR_i = \frac{\Sigma_{ii}^2}{\sum_i \Sigma_{ii}^2} = 0.29, 0.23, 0.17, 0.15, 0.15, 0.001$$

Question

What does this tell us?

Answer

One projected component contains only 0.1% of the data variability

Reducing dimensions

The 6th principal component looks like

$$\boldsymbol{V}_6 = \begin{pmatrix} 0.61 & 0.52 & 0.60 & -0.00 & -0.00 & -0.00 \end{pmatrix}^T$$

Question

Why do the entries look this way?

Reducing dimensions

The 6th principal component looks like

$$\boldsymbol{V}_6 = \begin{pmatrix} 0.61 & 0.52 & 0.60 & -0.00 & -0.00 & -0.00 \end{pmatrix}^T$$

Question

Why do the entries look this way?

Answer

Approximation of $M_{11} + M_{22} + M_{33} = 0$

Some remarks

- PCA is sensititive to scale of data
- Assumes linear relationship, cannot deal with time-shift, for example
- SVD/PCA is computationally expensive, specialized methods for large datasets

Kernel PCA

Similar to SVM we can expand PCA to non-linear cases using the kernel trick.

Linear: $k(\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$

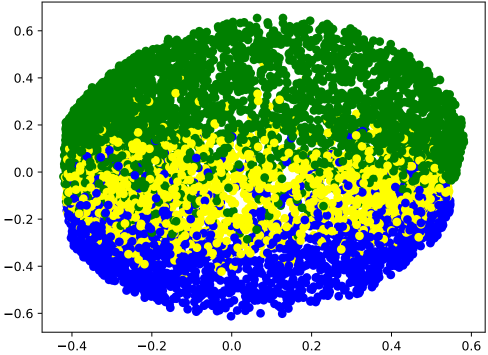
Polynomial: $k(\mathbf{a}, \mathbf{b}) = (\gamma \mathbf{a} \cdot \mathbf{b} + r)^d$

Gaussian RBF: $k(\mathbf{a}, \mathbf{b}) = \exp\left(-\gamma \|\mathbf{a} - \mathbf{b}\|\right)$

Sigmoid: $k(\mathbf{a}, \mathbf{b}) = \tanh\left(\gamma \mathbf{a} \cdot \mathbf{b} + r\right)$

Implicitly performs PCA in feature space of function Φ associated with kernel.

KPCA for moment tensor



Summary

- PCA is a useful tool for data analysis and dimensionality reduction
- Finds linear combinations that maximize/minimize variance
- Can analyze contributions to signal
- Can remove components with small contributions to make dataset more compact/ easier to visualize
- Kernel PCA is a non-linear extension to PCA