Week 14 Dimensionality Reduction

L14-01 Principal Component Analysis: Concept

Dong Hee Lee
Master's Student

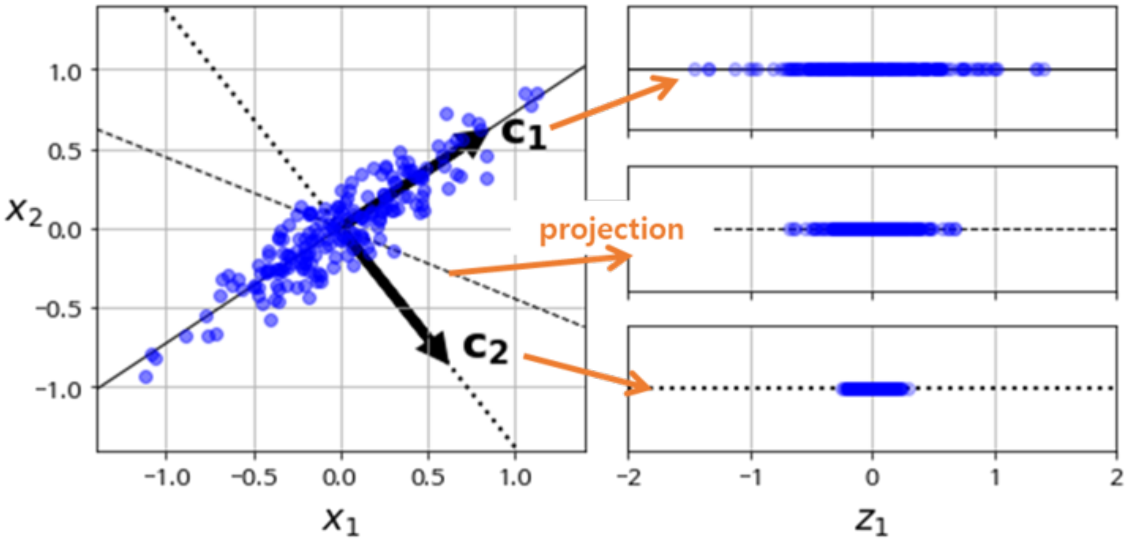
PCA Assumptions

- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality

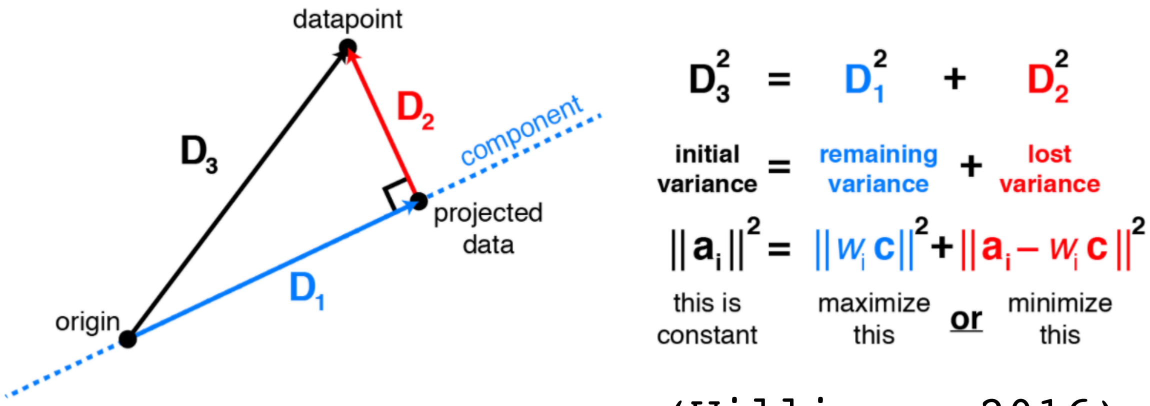
(Lee, 2021)

PCA Assumptions

- **Assumption #1**: Variance = relevance



(Géron, 2019)



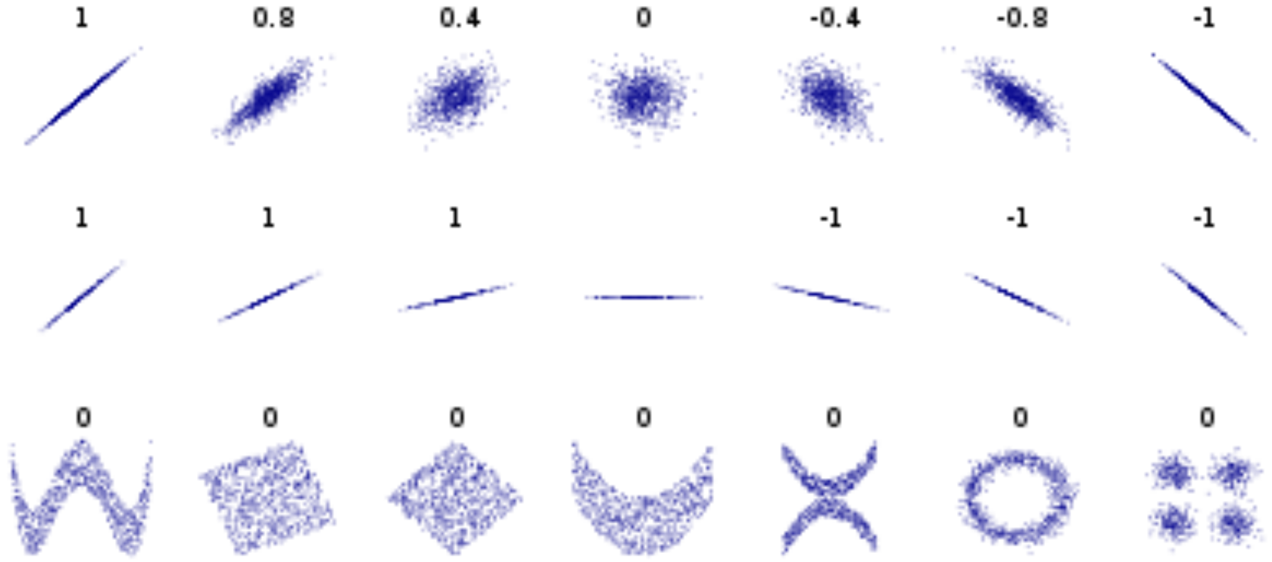
(Williams, 2016)

- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality

(Lee, 2021)

PCA Assumptions

- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
  - A PCA is based on Pearson correlation coefficients



- **Assumption #3**: Orthogonality

(Lee, 2021)

PCA Assumptions

- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality

https://mathinsight.org/applet/vector_3d_coordinates

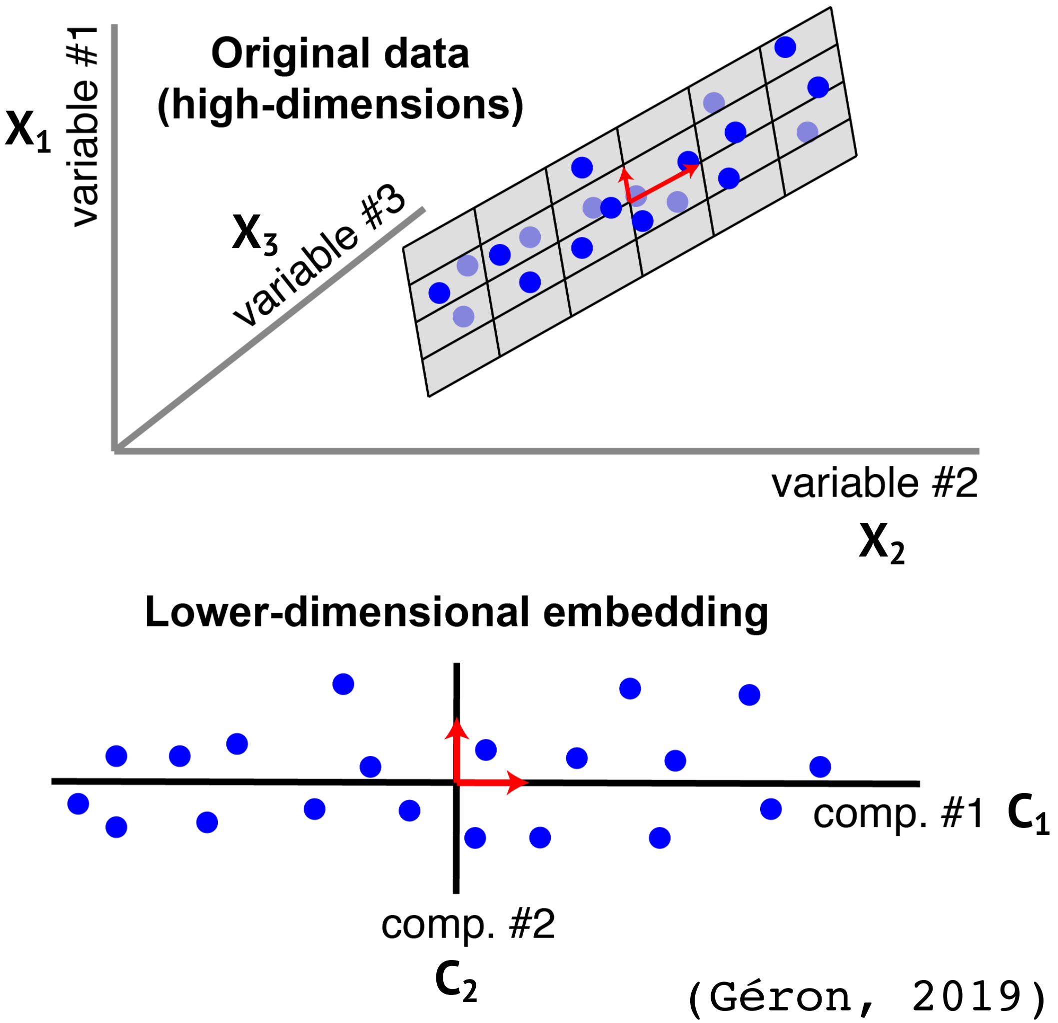
(Lee, 2021)

The Idea of PCA

■ **Four Steps**

1. PCA identifies the axis that provides **the largest amount of variance** in the dataset.
2. Then, it finds the second axis that (1) is **orthogonal with the first axis** and (2) has **the largest variance**.
3. Also, it finds the third axis that (1) is **orthogonal** to the first and second axes and (2) **preserves the variance** as much as possible.
4. In the same way as steps above, it **keeps finding the axes** as much as **the number of dimensions of the dataset**.

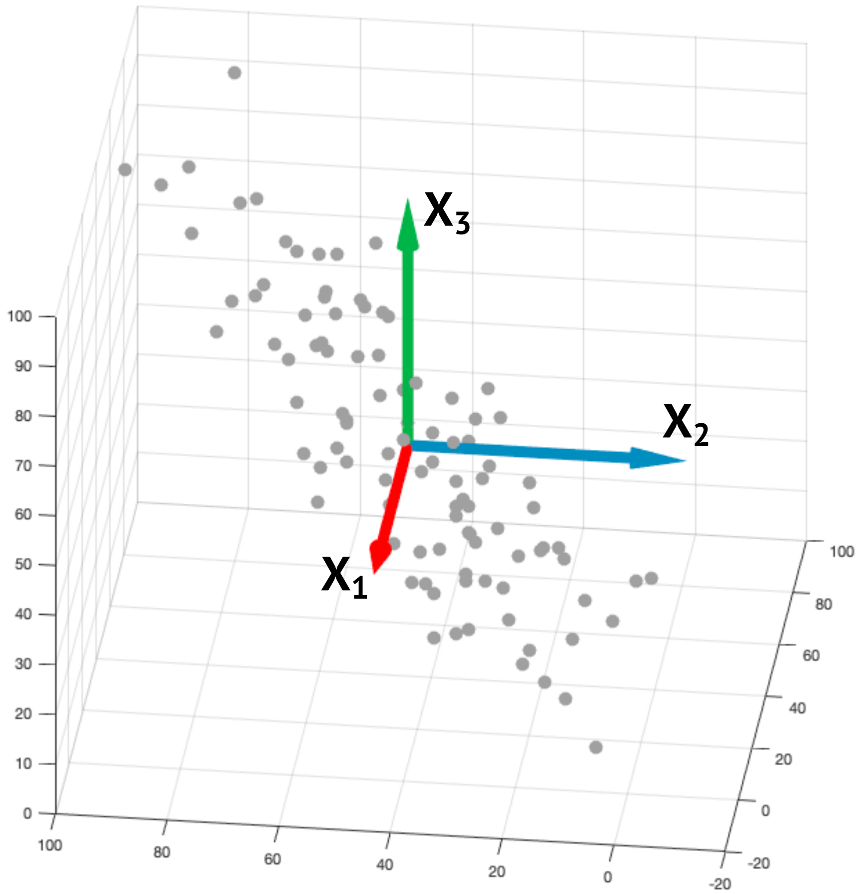
The Idea of PCA



(Géron, 2019)

The Idea of PCA

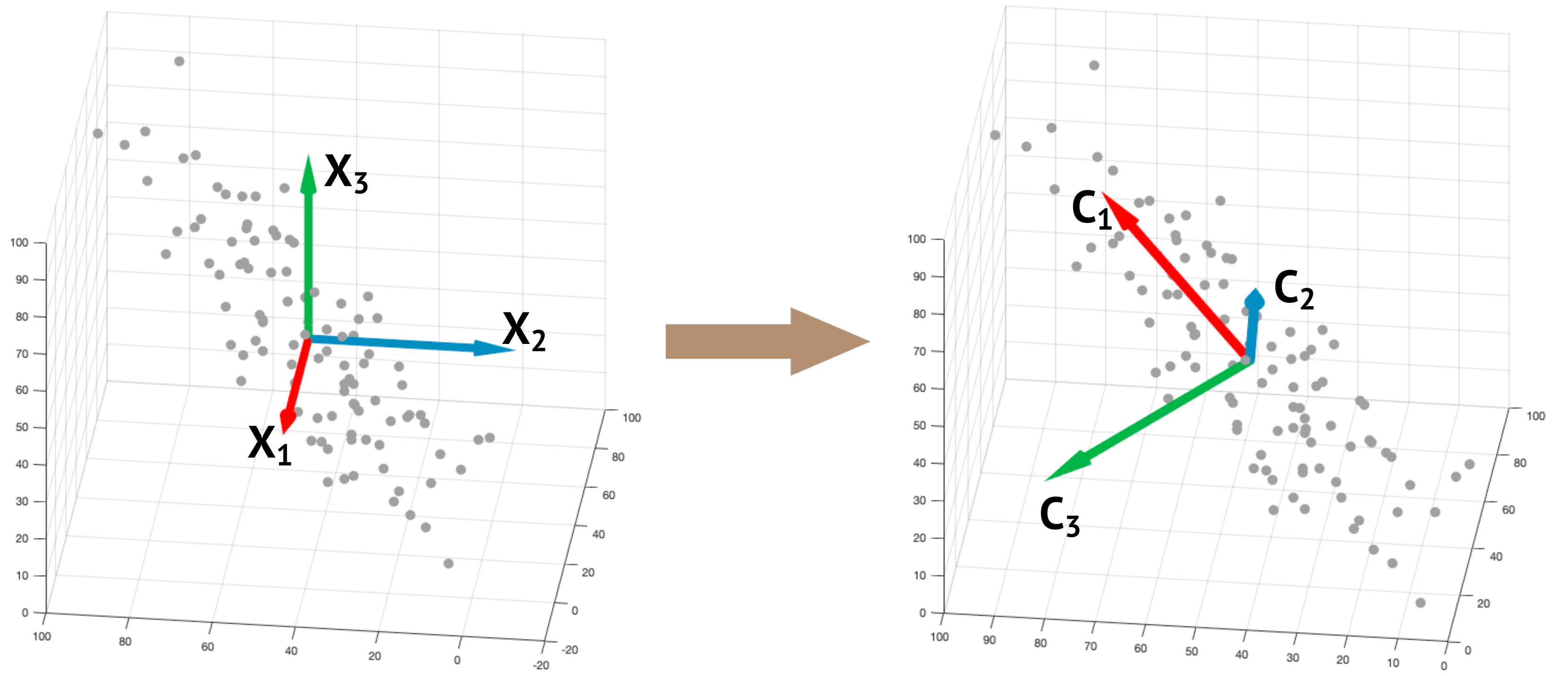
How can PCA identify the axis that provides the largest amount of variance in the dataset?



(Lee, 2021)

The Idea of PCA

How can PCA identify the axis that provides the largest amount of variance in the dataset?



(Lee, 2021)

The Idea of PCA

How can PCA identify the axis that provides the largest amount of variance in the dataset?

■ **Covariance:** A measure of the strength and sign of the linear relationship between two variables, in the scale of the original data

$$cov(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

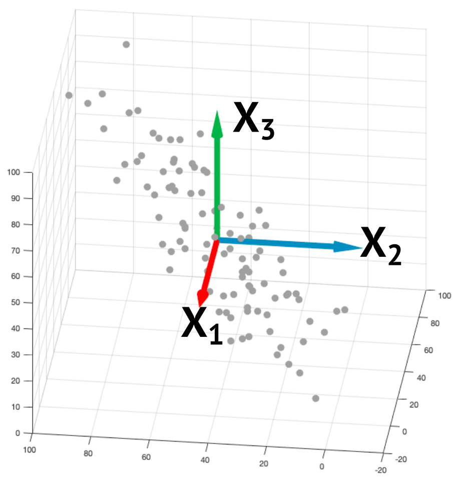
■ **Covariance matrix:** A square matrix giving the covariance between each pair of variables

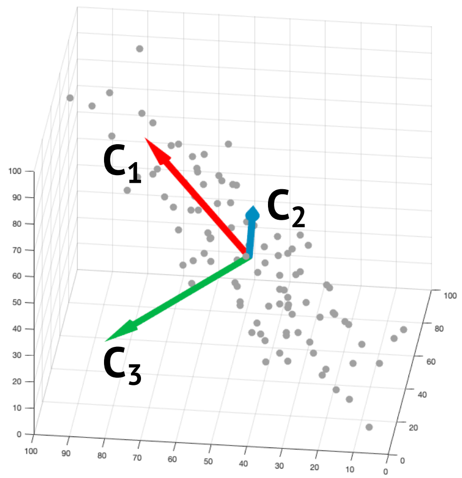
$$\Sigma = \begin{bmatrix} Var(X) & Cov(X, Y) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) & Cov(Y, Z) \\ Cov(Y, Z) & Cov(Y, Z) & Var(Z) \end{bmatrix}$$

(Lee, 2021)

The Idea of PCA

How can PCA identify the axis that provides the largest amount of variance in the dataset?





X Coordinate System (**X_1, X_2, X_3**)

$$\begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ Cov(X_1,X_2) & Var(X_2) & Cov(X_2,X_3) \\ Cov(X_1,X_3) & Cov(X_2,X_3) & Var(X_3) \end{bmatrix}$$

C Coordinate System (**C_1, C_2, C_3**)

$$\begin{bmatrix} Var(C_1) & 0 & 0 \\ 0 & Var(C_2) & 0 \\ 0 & 0 & Var(C_3) \end{bmatrix}$$

(Lee, 2021)

The Idea of PCA

How can PCA identify the axis that provides the largest amount of variance in the dataset?

$$A^T A$$

Covariance matrix
(Mean-centered, symmetric matrix)

$$(A = U\Sigma V^T)$$

Singular Value Decomposition of Data matrix

$$= (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T$$

eigenvectors

$$= \boxed{V}\boxed{(\Sigma\Sigma^T)}\boxed{V}^T$$

eigenvalues

Eigendecomposition of covariance matrix

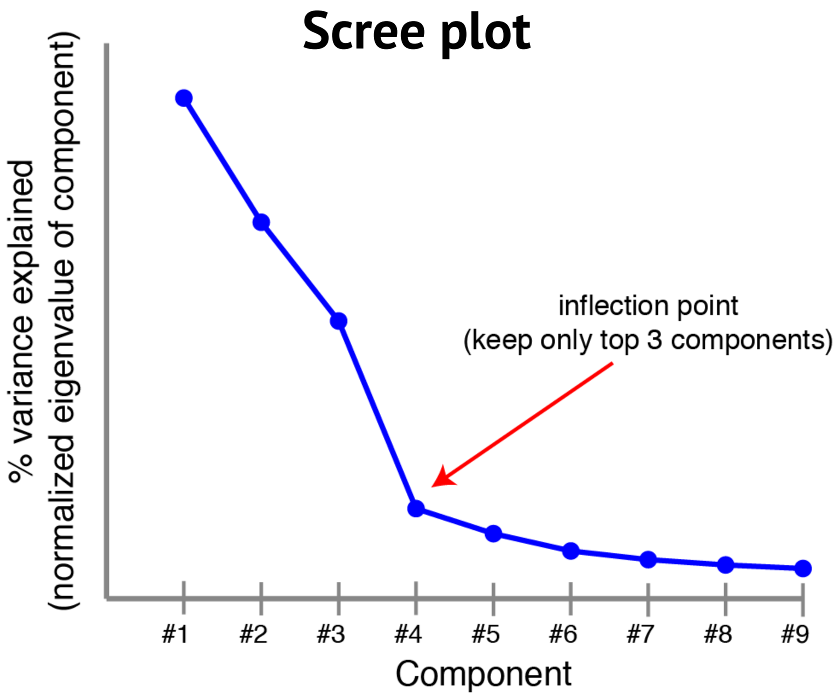
$$= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Principal component

Variance

The Idea of PCA

- **Choosing the number of components**



(Williams, 2016)

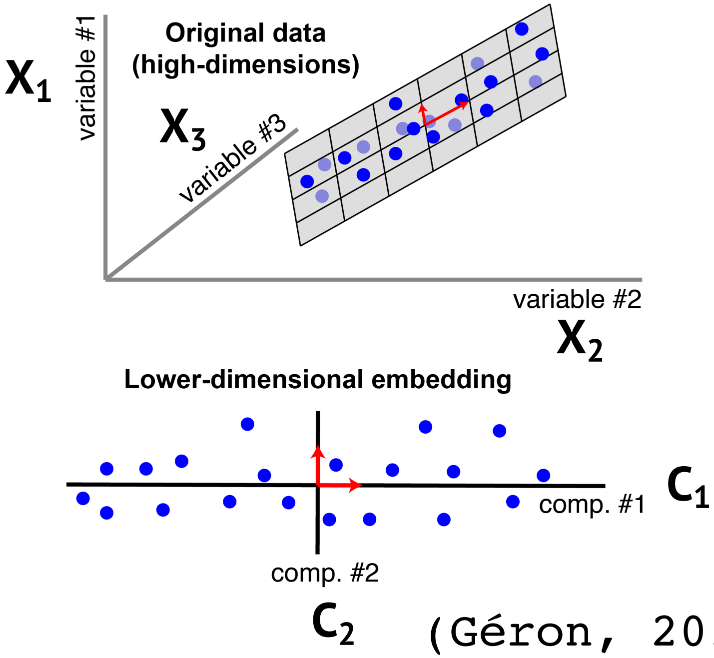
- Option #1: Elbow point

- Option #2: Select top K components that explain most variance in the data. (e.g., more than 70%)
  - $100 \times \frac{\sum_{j=1}^{K} \lambda_j}{\sum_{i=1}^{C} \lambda_i}$ (%)

Interim summary

■ **Assumptions**

- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality



(Géron, 2019)

■ How can PCA identify the axis that provides the largest amount of variance in the dataset?

- Measures how each variable is associated with on another using a **Covariance matrix**
- Understand the directions of the spread of our data using **Eigenvectors**
- Bring out the relative importance of these directions using **Eigenvalues**

■ **How many components?**

- Elbow points
- Top K components that explain most variance in the data

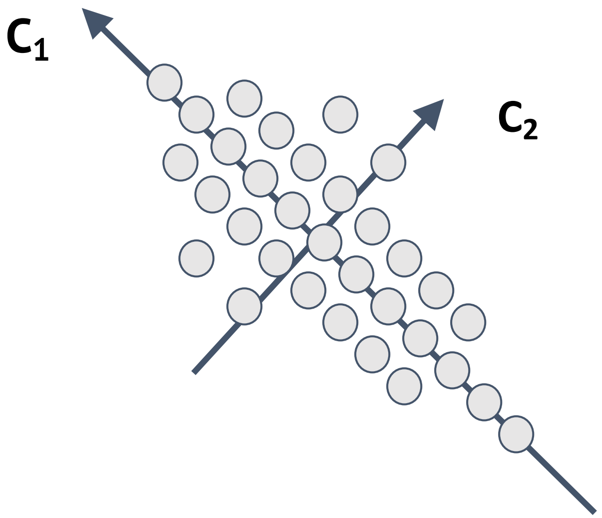
PCA Limitations

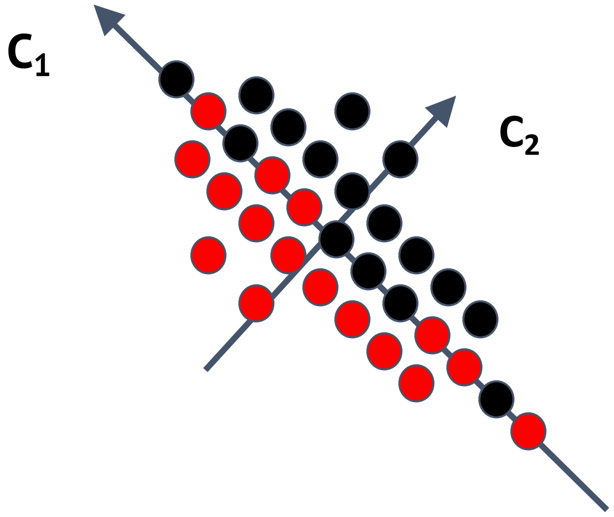
- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality

(Lee, 2021)

PCA Limitations

- **Assumption #1**: Variance = relevance





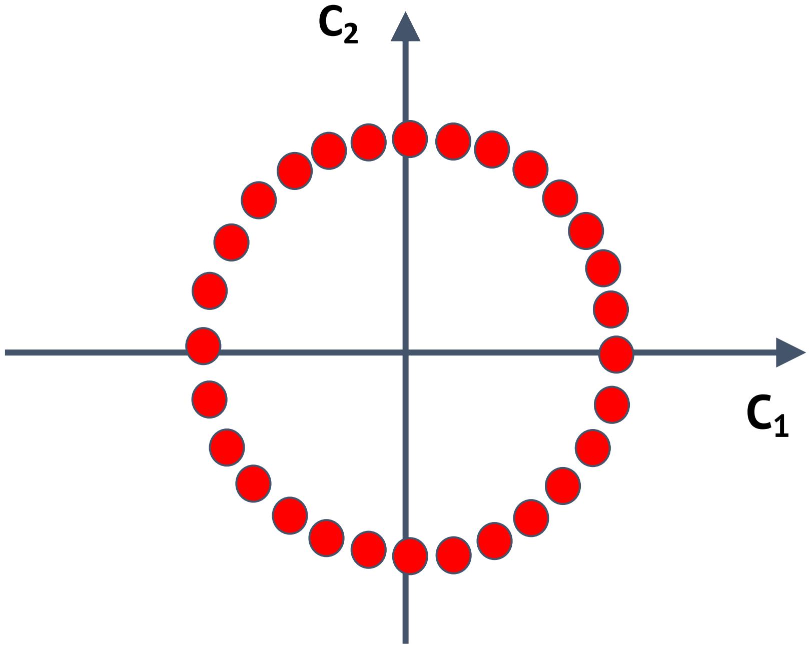
Use different algorithms for grouping
(e.g., K-means, Hierarchical clustering, Linear Discriminant Analysis, etc.)

- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality

(Lee, 2021)

PCA Limitations

- Assumption #1: Variance = relevance
- **Assumption #2**: Linear relationship



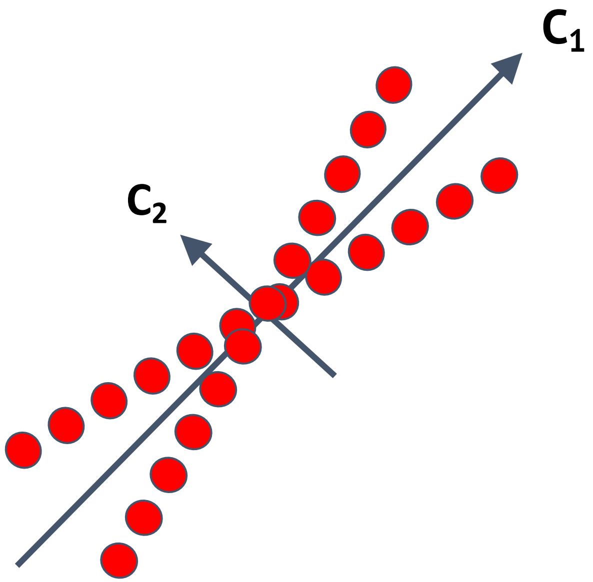
Use nonlinear algorithms
(e.g., Kernal PCA, multidimensional scaling, Laplacian eigenmap, autoencoder, etc.)

- Assumption #3: Orthogonality

(Lee, 2021)

PCA Limitations

- **Assumption #1**: Variance = relevance
- **Assumption #2**: Linear relationship
- **Assumption #3**: Orthogonality



Use different algorithms
(e.g., Independent component analysis, Nonlinear algorithms, Factor analysis)

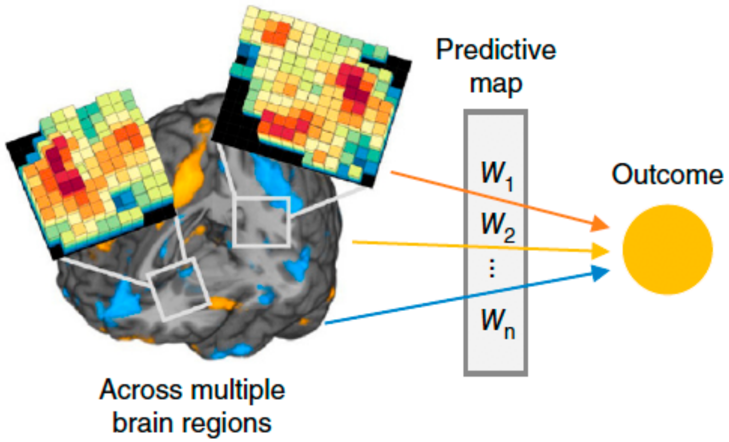
(Lee, 2021)

Application

- # of Features >> # of observations
- Multi-dimensional Data Interpretation
- Multicolinearlity
- Curse of dimensionality

Application

- **# of Features >> # of observations**



(Woo et al., 2017, *Nat Neurosci*)

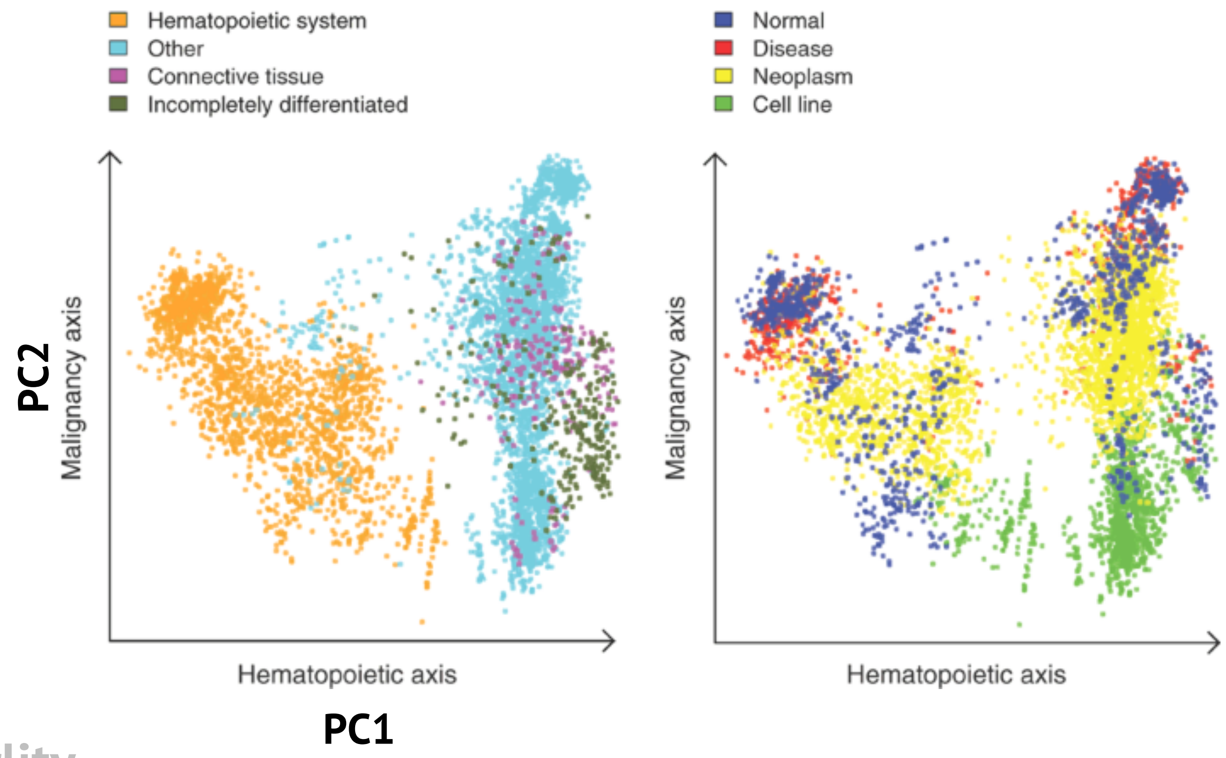
The prediction of pain rating based on Brain fMRI data

Features: > 200,000 fMRI voxels
Observations: < 200 samples

- Multi-dimensional Data Interpretation
- Multicolinearlity
- Curse of dimensionality

Application

- # of Features >> # of observations
- **Multi-dimensional Data Interpretation**



PCA for 14,000-dimension gene expression space across 5,372 samples

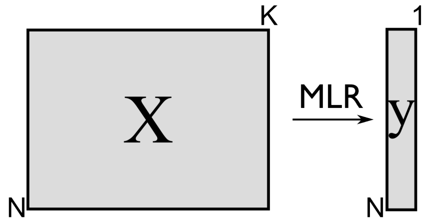
(Lukk et al., 2010, *Nat Biotechnol*)

- Multicolinearlity
- Curse of dimensionality

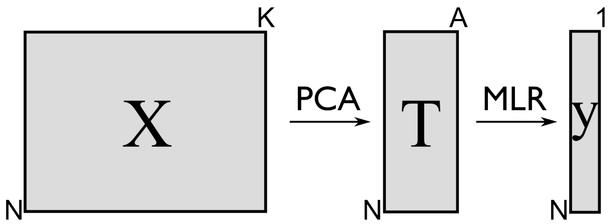
Application

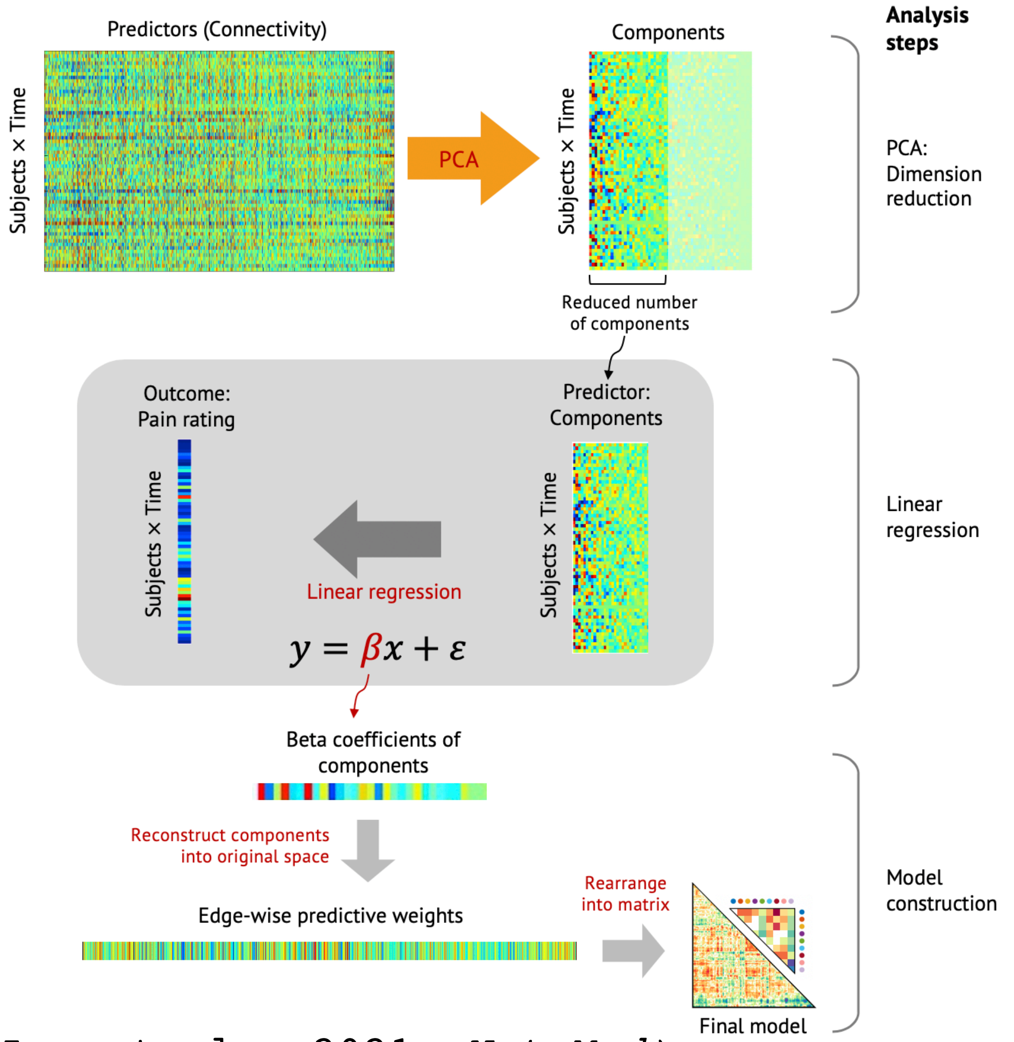
- # of Features >> # of observations
- Multi-dimensional Data Interpretation
- **Multicolinearlity**

Multiple linear regression



Principal component regression



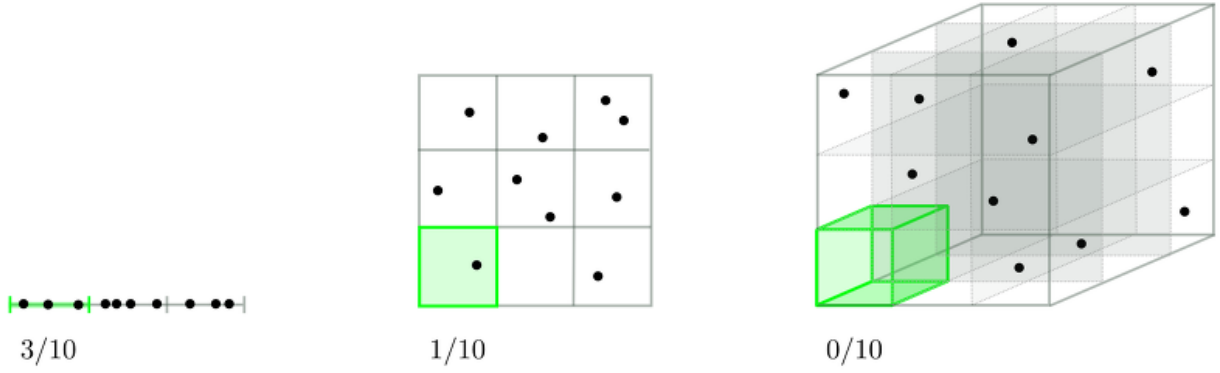


(Lee et al., 2021, *Nat Med*)

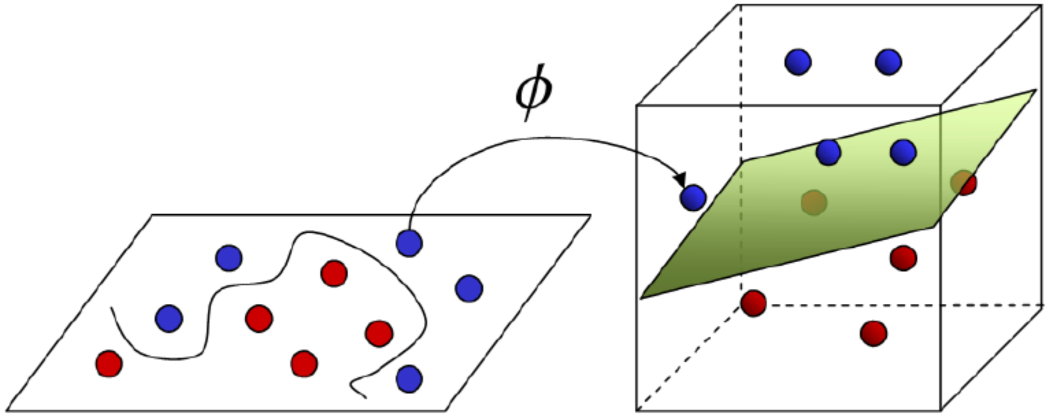
- Curse of dimensionality

Application

- # of Features >> # of observations
- Multi-dimensional Data Interpretation
- Multicolinearlity
- **Curse of dimensionality**



Curse of dimensionality



Blessing of dimensionality

Conclusion

- PCA is great for reducing data from M to C dimensions, where $1<C<M$.

- PCA identifies the (orthogonal) axes that provide the largest amount of variance in the dataset.
  - Eigendecomposition of Covariance matrix

- Remember that if PCA assumptions are violated, the results are invalid.

- PCA is useful for problems like too many features and multicollinearity.

Reference

- Mike X Cohen. PCA & multivariate signal processing, applied to neural data. *Udemy*. https://www.udemy.com/course/dimension-reduction-and-source-separation-in-neuroscience/
- Lee, Jae-Joong. (2021). Principal Component Analysis [PowerPoint slides]. *Department of Biomedical Engineering, Sungkyunkwan University*.
- Huettel, S. A., Song, A. W., & McCarthy, G. (2009). Functional magnetic resonance imaging. *Sunderland, Mass: Sinauer Associates*.
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems (2nd ed.). *O'Reilly*.
- Williams, A. (2016, Mar 27). Everything you did and didn't know about PCA. *Alex Williams's blog*. http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/

Reference

- Lever, J., Krzywinski, M. & Altman, N. (2017). Principal component analysis. *Nat Methods*. 14, 641–642. https://doi.org/10.1038/nmeth.4346