

Inclass 19: Dimensional Reduction

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.ac.kr)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

Principal component analysis (PCA)

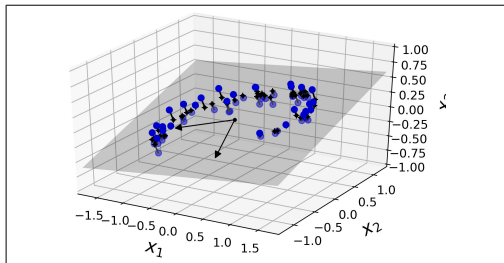


Figure 8-2. A 3D dataset lying close to a 2D subspace

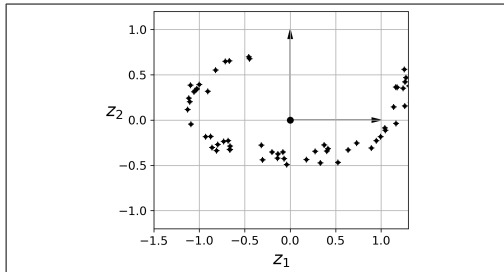


Figure 8-3. The new 2D dataset after projection

Principal component analysis (PCA)

Covariance measures the strength of the linear relationship between two variables

$$\sigma_{xy} = \text{E}[(x - \mu_x)(y - \mu_y)] \quad (1)$$

Covariance matrix **C** for multivariate random variable **x**

$$C_{ij} = \text{E}[(x_i - \mu_i)(x_j - \mu_j)] \quad (2)$$

Principal component analysis (PCA)

Preserving the variance

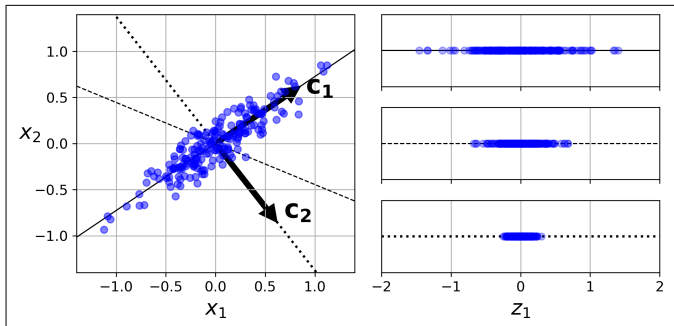


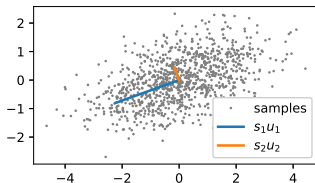
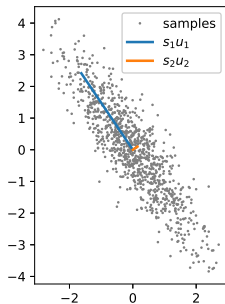
Figure 8-7. Selecting the subspace onto which to project

Principal component analysis (PCA)

For given data $x_1, x_2, \dots, x_N \in \mathbb{R}^D$

1. create a matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ with one column vector per each sample
2. covariance matrix $\mathbf{C} = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T] \in \mathbb{R}^{D \times D}$
3. find singular vectors and singular values of $\mathbf{C}$
4. principal components = largest singular values and vectors

Principal component analysis (PCA)



Matrix factorization

- Principal component analysis (PCA): orthogonal property
- Vector quantization (VQ): unary property
- Non-negative matrix factorization (NMF): non-negativity

PCA: $\mathbf{X} \approx \mathbf{WH}$

where $\mathbf{W}^T \mathbf{W} = \mathbf{I}$

VQ: $\mathbf{X} \approx \mathbf{WH}$

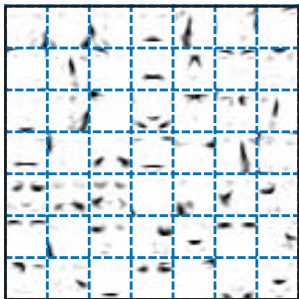
where $\mathbf{H}$ consists of unary vectors

NMF: $\mathbf{X} \approx \mathbf{WH}$

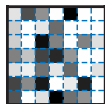
where $W_{ij} \geq 0, H_{ij} \geq 0$

Matrix factorization and face image compression

NMF



$\times$



$=$



Original



VQ



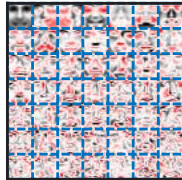
$\times$



$\parallel$



PCA



$\times$



$\parallel$



Nonnegative matrix factorization (NMF)

Nonnegative matrix factorization (NMF)

given $\mathbf{X} \in \mathbb{R}^{n \times m}$

find $(\mathbf{W}, \mathbf{H}) = \arg \min_{(\mathbf{W}, \mathbf{H})} \|\mathbf{X} - \mathbf{WH}\|_F^2$

s.t. $W \in \mathbb{R}_+^{n \times k}$ and $H \in \mathbb{R}_+^{k \times m}$ ($k \leq \text{rank}(A)$)

PCA vs. NMF for MNIST dataset

