PRINCIPAL COMPONENT ANALYSIS

CS6140

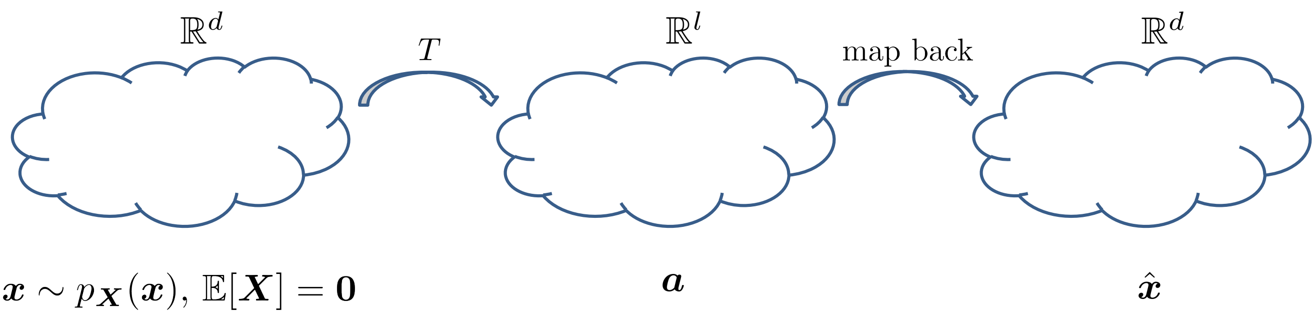
Predrag Radivojac
KHOURY COLLEGE OF COMPUTER SCIENCES
NORTHEASTERN UNIVERSITY

Spring 2021

Problem Formulation

Given: a set of vectors $\{\boldsymbol{x}_i\}_{i=1}^{n}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$, sampled from $p_{\boldsymbol{X}}(\boldsymbol{x})$

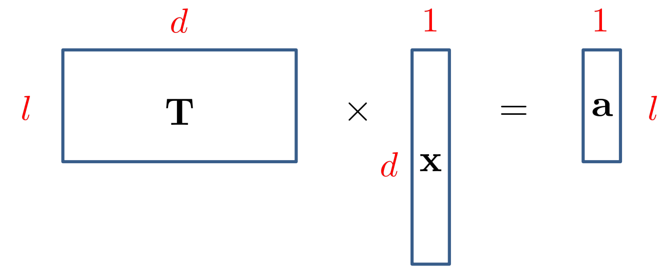
Objective: find a linear mapping $T : \mathbb{R}^d \to \mathbb{R}^l$, where $l < d$, such that the reconstruction of projections back to $\mathbb{R}^d$ is optimal in the mean-squared-error sense.



Problem Formulation

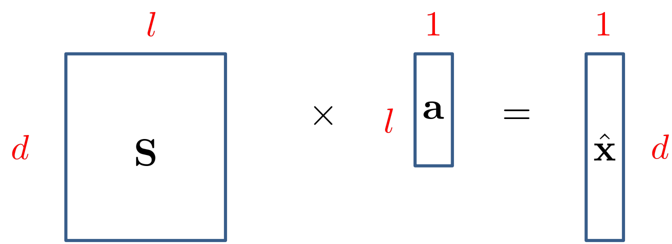
Matrix view: $\mathbf{x} \in \mathbb{R}^{d\times 1}$, $\mathbf{T} \in \mathbb{R}^{l\times d}$. The goal is to find $\mathbf{T}$.

$\mathbf{Tx} = \mathbf{a}$

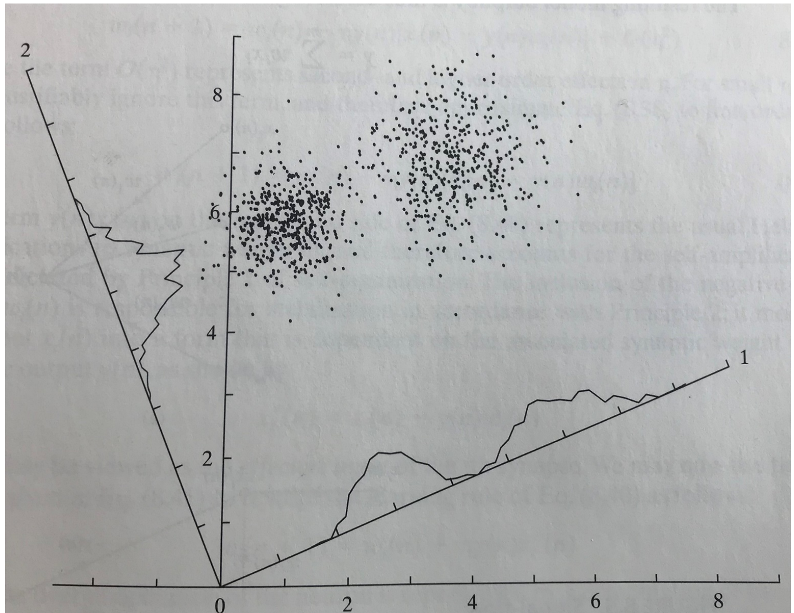


Minimize: $\mathbb{E}[||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2]$

$\mathbf{Sa} = \hat{\mathbf{x}}$



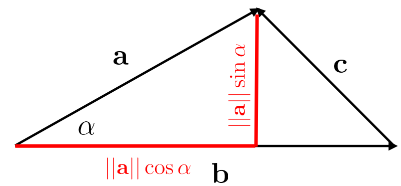
Idea



Haykin. Neural networks. 1999.

Preliminaries: Proof for Cosine

$$\mathbf{a} = \mathbf{b} + \mathbf{c}$$



$$\begin{aligned}||\mathbf{c}||^2 &= (||\mathbf{b}|| - ||\mathbf{a}||\cos\alpha)^2 + (||\mathbf{a}||\sin\alpha)^2 \\ &= ||\mathbf{b}||^2 - 2||\mathbf{a}||\cdot||\mathbf{b}||\cos\alpha + ||\mathbf{a}||^2\cos^2\alpha + ||\mathbf{a}||^2\sin^2\alpha \\ &= ||\mathbf{a}||^2 + ||\mathbf{b}||^2 - 2||\mathbf{a}||\cdot||\mathbf{b}||\cos\alpha\end{aligned}$$

$$\cos(\alpha) = \frac{\mathbf{a}^T\mathbf{b}}{||\mathbf{a}||\cdot||\mathbf{b}||}$$

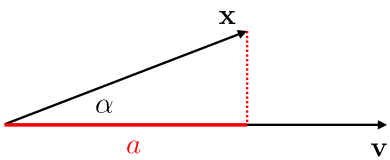
$$\begin{aligned}||\mathbf{c}||^2 &= \mathbf{c}^T\mathbf{c} \\ &= (\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b}) \\ &= \mathbf{a}^T\mathbf{a} - 2\mathbf{a}^T\mathbf{b} + \mathbf{b}^T\mathbf{b} \\ &= ||\mathbf{a}||^2 - 2\mathbf{a}^T\mathbf{b} + ||\mathbf{b}||^2\end{aligned}$$

Combine the two:

$$||\mathbf{a}||^2 - 2\mathbf{a}^T\mathbf{b} + ||\mathbf{b}||^2 = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 - 2||\mathbf{a}||\cdot||\mathbf{b}||\cos\alpha$$

Projection to One Dimension

Let us project a vector $\mathbf{x}$ to a unit vector $\mathbf{v}$. Note: $\mathbf{v}^T\mathbf{v} = 1$ or $||\mathbf{v}|| = 1$.



$$\cos(\alpha) = \frac{a}{||\mathbf{x}||} = \frac{\mathbf{x}^T\mathbf{v}}{||\mathbf{x}||\cdot||\mathbf{v}||} \qquad \Rightarrow \qquad a = \mathbf{x}^T\mathbf{v}$$

Let us project a random vector $\overset{d\times 1}{\boldsymbol{X}} \sim p(\boldsymbol{x})$ to some unit vector $\mathbf{v}$.

$$A = \boldsymbol{X}^T\mathbf{v} = \mathbf{v}^T\boldsymbol{X}$$

$$\mathbb{E}[A] = \mathbf{v}^T\mathbb{E}[\boldsymbol{X}] = 0$$

$$\mathbb{E}[A^2] = \mathbb{E}[\mathbf{v}^T\boldsymbol{X}\boldsymbol{X}^T\mathbf{v}] = \mathbf{v}^T\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]\mathbf{v} = \mathbf{v}^T\overset{\overset{d\times d}{\downarrow}}{\Sigma}\mathbf{v} = \mathbb{V}[A]$$

Projection to One Dimension

For a set of vectors, let us find a unit vector $\mathbf{v}$ so that the projection has maximum variance $\mathbb{V}[A] = \mathbf{v}^T \Sigma \mathbf{v}$.

Objective: Maximize variance of the projection.

$$\max \mathbf{v}^T \Sigma \mathbf{v} \quad \text{s.t. } \mathbf{v}^T \mathbf{v} = 1$$

$$L(\mathbf{v}, \lambda) = \mathbf{v}^T \Sigma \mathbf{v} + \lambda(1 - \mathbf{v}^T \mathbf{v}) \qquad \xRightarrow{\text{Solve}} \qquad \Sigma \mathbf{v} = \lambda \mathbf{v}$$

The eigenvalue problem

PROJECTION TO *d* DIMENSIONS

Consider now projecting to d orthogonal vectors:

$$\mathbf{\Sigma V} = \mathbf{V\Lambda}$$

The eigenvalue problem

where $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_d]$, with $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_d\}, \text{ with } \lambda_1 \geq \lambda_2 \ldots \geq \lambda_d$$

Let us re-write: $\mathbf{V}^T\mathbf{\Sigma V} = \mathbf{\Lambda}$

$$\mathbf{v}_i^T\mathbf{\Sigma}\mathbf{v}_j = \begin{cases} \lambda_i & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

$\leftarrow$ variance of projection A_i

Transformation

Let us express the i-th projection as $a_i = \mathbf{v}_i^T \mathbf{x} = \mathbf{x}^T \mathbf{v}_i$

Thus,

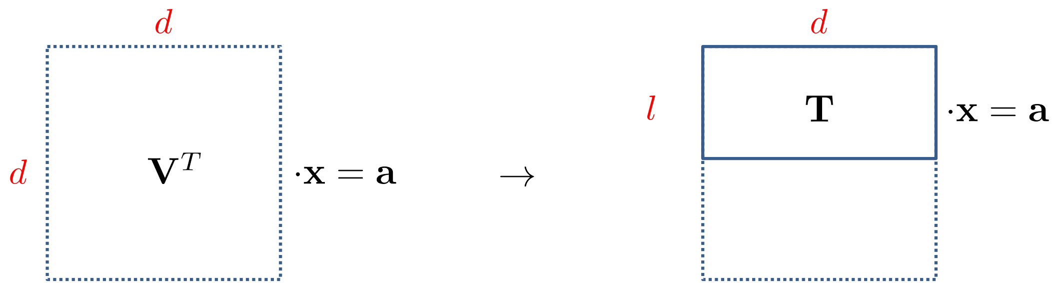
$$\mathbf{a} = (a_1, a_2, \ldots, a_d) = (\mathbf{v}_1^T \mathbf{x}, \mathbf{v}_2^T \mathbf{x}, \ldots, \mathbf{v}_d^T \mathbf{x}) = \mathbf{V}^T \mathbf{x} = \sum_{i=1}^{d} x_i \mathbf{v}_i^T$$

Let us reconstruct $\mathbf{x}$ now. Remember, $\mathbf{V}^{-1} = \mathbf{V}^T$.

$$\mathbf{x} = \mathbf{Va} = \sum_{i=1}^{d} a_i \mathbf{v}_i$$

Dimensionality Reduction

Let us try to keep the first l components of $\mathbf{a}$.



Let us reconstruct $\mathbf{x}$ now

$$\hat{\mathbf{x}} = \sum_{i=1}^{l} a_i \mathbf{v}_i$$

Dimensionality Reduction

The error vector $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is now

$$\mathbf{e} = \sum_{i=l+1}^{d} a_i \mathbf{v}_i$$

Because

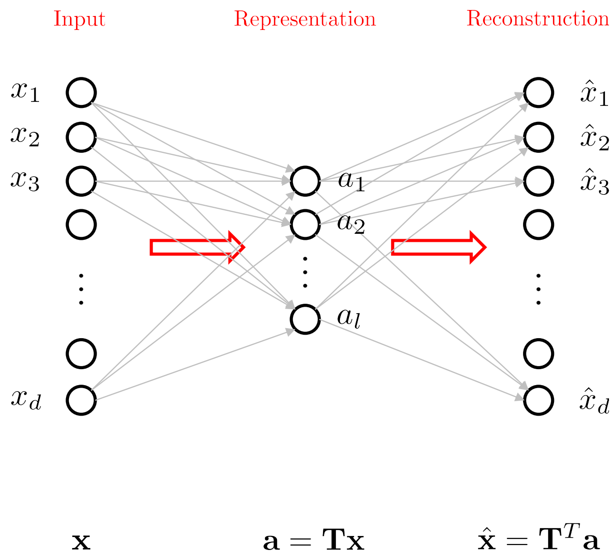
$$\mathbf{x} - \hat{\mathbf{x}} = \sum_{i=1}^{d} a_i \mathbf{v}_i - \sum_{i=1}^{l} a_i \mathbf{v}_i$$

We now have

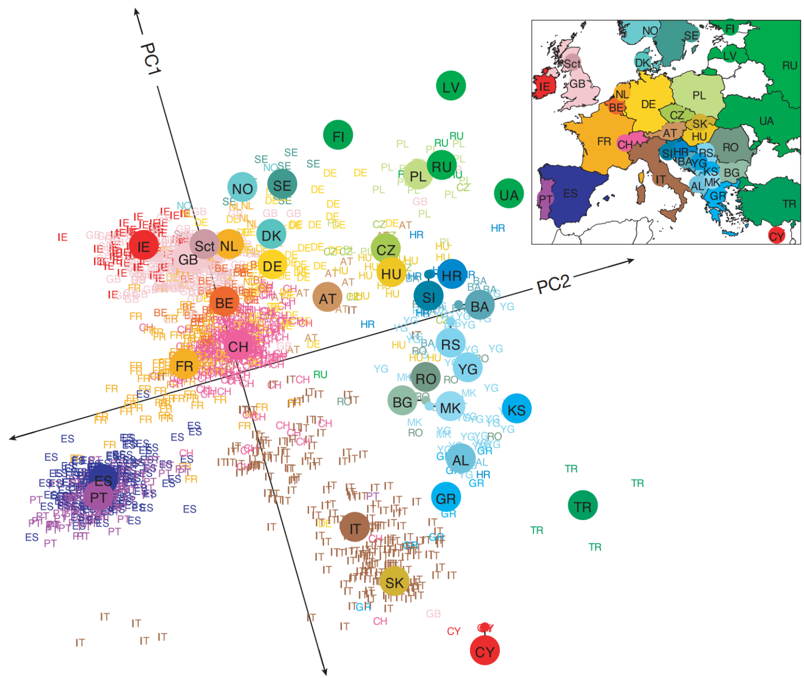
$$\mathbb{E}[\boldsymbol{X} - \hat{\boldsymbol{X}}] = \mathbf{0} - \textstyle\sum_{i=1}^{l} \mathbb{E}[A_i]\mathbf{v}_i = \mathbf{0}$$

$$\mathbb{E}[||\boldsymbol{X} - \hat{\boldsymbol{X}}||^2] = \textstyle\sum_{i=l+1}^{d} \mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_i = \sum_{i=l+1}^{d} \lambda_i \qquad \leftarrow \text{proved later}$$

Different Type of Visualization



VISUALIZATION



Novembre et al. Genes mirror geography within Europe. *Nature*, 2008.

Relationship with Singular Value Decomposition (SVD)

Every matrix $\mathbf{X}$ has a SVD: $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

$\mathbf{U}$ = orthogonal
$\mathbf{S}$ = diagonal
$\mathbf{V}^T$ = orthogonal

In MATLAB: $[\mathrm{U}, \mathrm{S}, \mathrm{V}] = \mathrm{svd}(\mathrm{X})$

Let's look at $\mathbf{X}^T\mathbf{X}$

$$\mathbf{X}^T\mathbf{X} = (\mathbf{U}\mathbf{S}\mathbf{V}^T)^T(\mathbf{U}\mathbf{S}\mathbf{V}^T) = \mathbf{V}\mathbf{S}^T\mathbf{S}\mathbf{V}^T = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T.$$

$$\boldsymbol{\Sigma} = \mathbf{S}^T\mathbf{S}.$$

Appendix

Proof for the squared norm of the error vector:

$$
\begin{aligned}
\mathbb{E}[||(\boldsymbol{X} - \hat{\boldsymbol{X}})||^2] &= \mathbb{E}[(\boldsymbol{X} - \hat{\boldsymbol{X}})^T(\boldsymbol{X} - \hat{\boldsymbol{X}})] \\
&= \mathbb{E}[\boldsymbol{X}^T\boldsymbol{X}] - 2\mathbb{E}[\boldsymbol{X}^T\hat{\boldsymbol{X}}] + \mathbb{E}[\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}]
\end{aligned}
$$

We investigate one of these terms

$$
\mathbb{E}[\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}}] = \mathbb{E}[\sum_{i=1}^{l} A_i\mathbf{v}_i^T \cdot \sum_{j=1}^{l} A_j\mathbf{v}_j] = \mathbb{E}[\sum_{i=1}^{l} A_i^2\mathbf{v}_i^T\mathbf{v}_i] = \mathbb{E}[\sum_{i=1}^{l} A_i^2]
$$

because $\mathbf{v}_i^T\mathbf{v}_j = 0$ when $i \neq j$ and $\mathbf{v}_i^T\mathbf{v}_j = 1$ when $i = j$. This makes a double sum above a single sum.

Appendix

We similarly have

$$\mathbb{E}[\boldsymbol{X}^T\boldsymbol{X}] = \mathbb{E}[\sum_{i=1}^{d} A_i \mathbf{v}_i^T \cdot \sum_{j=1}^{d} A_j \mathbf{v}_j] = \mathbb{E}[\sum_{i=1}^{d} A_i^2]$$

$$\mathbb{E}[\boldsymbol{X}^T\hat{\boldsymbol{X}}] = \mathbb{E}[\sum_{i=1}^{d} A_i \mathbf{v}_i^T \cdot \sum_{j=1}^{l} A_j \mathbf{v}_j] = \mathbb{E}[\sum_{i=1}^{l} A_i^2]$$

Finally, we have

$$\mathbb{E}[(\boldsymbol{X}-\hat{\boldsymbol{X}})^T(\boldsymbol{X}-\hat{\boldsymbol{X}})] = \sum_{i=1}^{d} \lambda_i - 2\sum_{i=1}^{l} \lambda_i + \sum_{i=1}^{l} \lambda_i = \sum_{i=l+1}^{d} \lambda_i$$

Q.E.D.