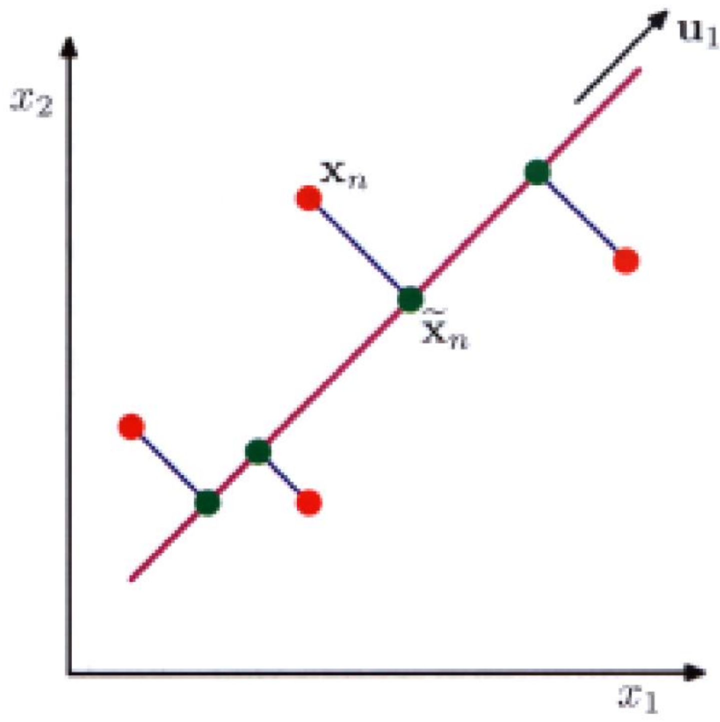
A quick view of Principal Component Analysis

Alberto Paccanaro
EMAp – FGV

www.paccanarolab.org

What we want



A space of lower dimensionality such that the orthogonal projection of the datapoints:

1) **maximizes the variance** of the projected points

2) **minimizes the sum of square of the projection errors**

The math

$\{\boldsymbol{x}_n\} \quad n = 1 \ldots N$, points in D dimensional space (column)

We want to project them into $M<D$ dimensions

Consider $M = 1$, defined by $\boldsymbol{u}_1$ unit vector (column).

Each point is projected into a scalar $\boldsymbol{u}_1^T \boldsymbol{x}_n$

$\overline{\boldsymbol{x}}$ mean of projected data:

$$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n$$

variance of projected data:

$$\frac{1}{N} \sum_{n=1}^{N} \left\{ \mathbf{u}_1^{\mathrm{T}} \mathbf{x}_n - \mathbf{u}_1^{\mathrm{T}} \overline{\mathbf{x}} \right\}^2 = \mathbf{u}_1^{\mathrm{T}} \mathbf{S} \mathbf{u}_1$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}$$

We want to maximize the projected variance $\boldsymbol{u}_1^T \boldsymbol{S}\, \boldsymbol{u}_1$ with respect to $\boldsymbol{u}_1$.

To enforce $\boldsymbol{u}_1^T \boldsymbol{u}_1 = 1$ we use a Lagrange multiplier, λ_1, and maximize:

$$\mathbf{u}_1^{\mathrm{T}} \mathbf{S} \mathbf{u}_1 + \lambda_1 \left(1 - \mathbf{u}_1^{\mathrm{T}} \mathbf{u}_1\right)$$

Setting the derivative w.r.t. $\boldsymbol{u}_1$ equal to zero:

$$\mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

So $\boldsymbol{u}_1$ must be an eigenvector of $\boldsymbol{S}$!

projected variance!

$$\mathbf{u}_1^{\mathrm{T}} \mathbf{S} \mathbf{u}_1 = \lambda_1$$

➔ the variance is maximized by setting $\boldsymbol{u}_1$ equal to the eigenvector having the largest eigenvalue

We can define additional components incrementally, choosing new directions to be the ones that maximize the projected variance among the directions orthogonal to those already considered…

To summarize…

The optimal linear projection for which the variance of the projected data is maximized is defined by

the M eigenvectors $\boldsymbol{u}_1 \ldots \boldsymbol{u}_M$ of the data covariance matrix $\mathbf{S}$ corresponding to the M largest eigenvalues $\lambda_1 \ldots \lambda_M$

Computational complexity:

- for a DxD matrix, $O(D^3)$
- Top eigenvalues and eigenvectors, $O(M\,D^2)$

Important – don't forget the mean ☺

- The projection onto the Principal Components makes more sense once you have subtracted the mean from your data.

- But when you want to reconstruct your points in the original space you will need to:
    1. Get back onto the original coordinate system
    2. **Add the mean back !**