Bernd Hamann 4/9/2021

Stratovan

■ DIMENSION REDUCTION AND DATA REDUCTION — PRINCIPAL COMPONENT ANALYSIS (PCA), SINGULAR VALUE DECOMPOSITION (SVD), KARHUNEN-LOEVE TRANSFORMATION (KLT)

- **Motivation:**
  - → Sample/training data sets used for classification are very large and high-dimensional.
  - → When an original feature point/vector data set consists of data that lies on an inherent manifold/hypersurface of much smaller dimensionality than the original high-dimensional space used to represent the data, dimension reduction could greatly reduce dimensionality.
  - → Since dimension reduction uses many data points/vectors to define a "local coordinate system and manifold/hypersurface to which the data are close," data can be replaced by an explicit, compact representation of the manifold — leading to reduction of data size.
  - → Dimension reduction applied to a very large, high-dimensional feature point/vector data set can significantly speed up classification:
    (i) Data dimensionality is reduced.
    (ii) Data set size (number of data primitives) is reduced.
  - → The manifold/hypersurface used to approximate and represent a multitude of original feature points/vectors must define (i) Location and geometry of the manifold (via a coordinate system) and (ii) the boundary of the manifold.

DIMENSION REDUCTION — PCA, SVD, KLT — Cont'd.

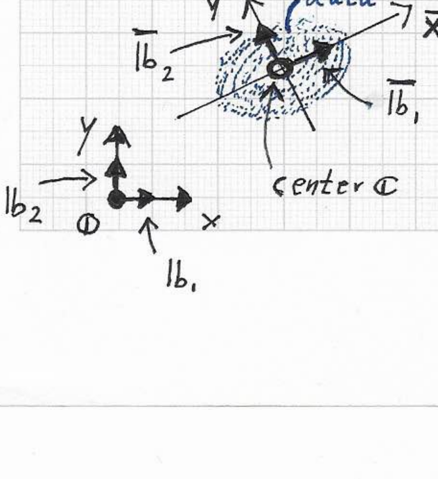
[Using only the term PCA in the following]

- Issue: "More complex feature data require more complex application of the basic PCA approach." One can consider these cases:

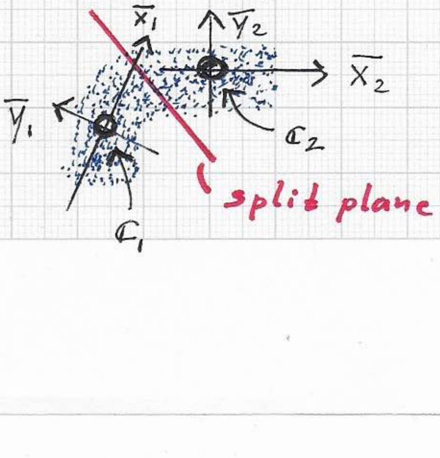
  1) One performs basic PCA for the entire data set, and one PCA step suffices.
  2) One must perform the basic PCA step many times (iteratively or recursively) for many data subsets defined via data splitting.
  3) One can/must a hierarchical, multi-resolution representation of PCA to support multiple levels of data approximation quality or to support progressive data classification.
  4) One can consider the use of weighted PCA — to assign higher weights to certain original data or to assign a higher weight to a datum that represents a cluster of original data — when dealing with noise, uncertainty etc.

---

1) Basic PCA:



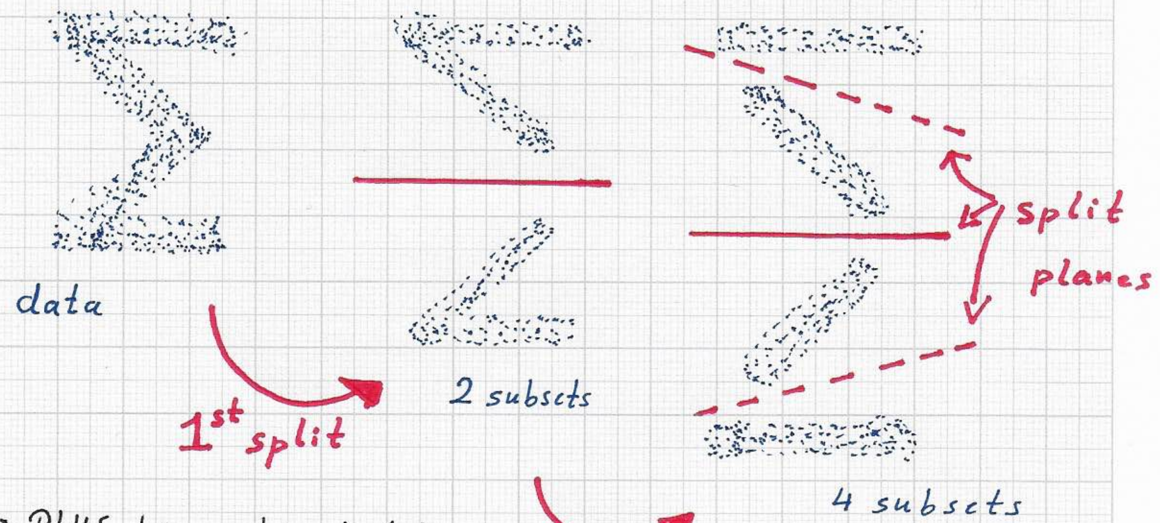
2) Splitting + PCA:



⇒ must use near-optimal, efficient splitting method

DIMENSION REDUCTION — Cont'd.

... 3) Multi-resolution PCA: Apply splitting recursively

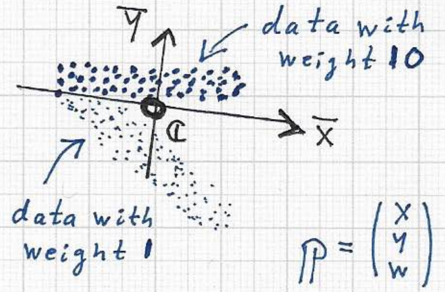


→ PLUS hierarchical data structure to support classification at increasing degree of confidence

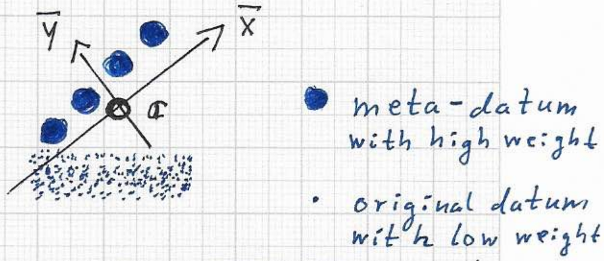
4) Weighted PCA: (i) original data with weights (e.g., noise-based)

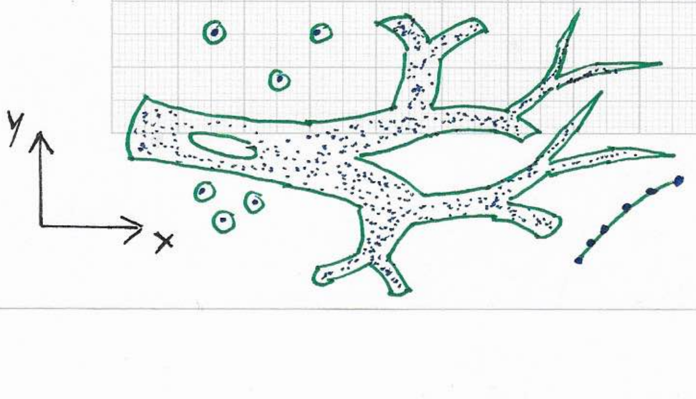
(ii) higher weights for derived "meta-data" representing clusters of original data

(i) weighted data:



(ii) "meta-data" = cluster representatives:

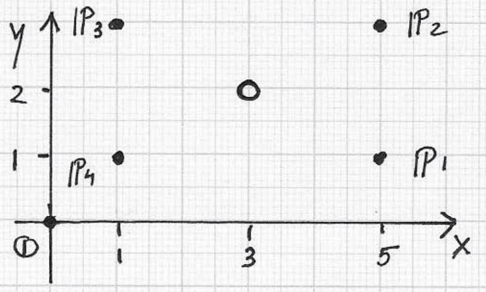




• Note: Illustration shows a "pathological case", where (feature) data reside in 2D space and locally define 0D, 1D and 2D manifold regions.

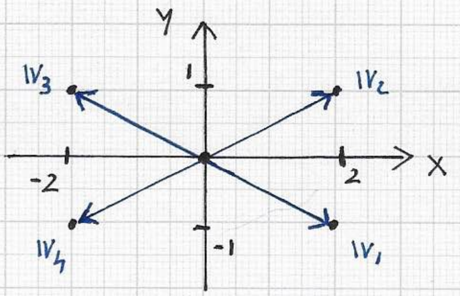
DIMENSION REDUCTION – Cont'd.

- **Ex.:** 2D example demonstrating essence of PCA



- given: $\mathbb{P}_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$, $\mathbb{P}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$, $\mathbb{P}_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\mathbb{P}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- center $\mathbb{C} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ ("mean")
- perform mean-subtraction:

$$\vec{\mathbb{V}}_i = \mathbb{P}_i - \mathbb{C},$$



$$\mathbb{V}_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix},\ \mathbb{V}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix},\ \mathbb{V}_3 = \begin{pmatrix} -2 \\ 1 \end{pmatrix},\ \mathbb{V}_4 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

($\mathbb{V}_i = \vec{\mathbb{V}}_i$ are positional VECTORS.)

- determine eigenvalues and eigenvectors:

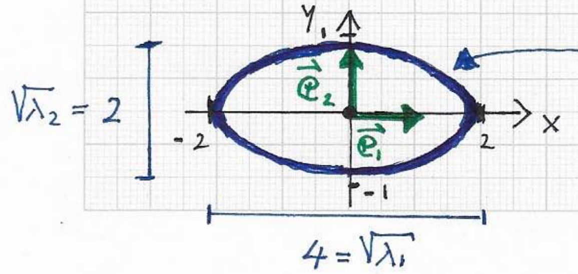
$$\mathbf{C} = \begin{pmatrix} 2 & 2 & -2 & -2 \\ -1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 2 & 1 \\ -2 & 1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}$$

- characteristic polynomial of $\mathbf{C}$: $p(\lambda) = \begin{vmatrix} 16-\lambda & 0 \\ 0 & 4-\lambda \end{vmatrix}$

$$= (16-\lambda)(4-\lambda)$$

- C's eigenvalues: $\lambda_1 = 16$, $\lambda_2 = 4$
- C's eigenvectors: $\begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \lambda_i \begin{pmatrix} x_i \\ y_i \end{pmatrix}$, $i = 1, 2$

$$\Rightarrow \vec{\mathbb{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \vec{\mathbb{e}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{(normalized)}$$



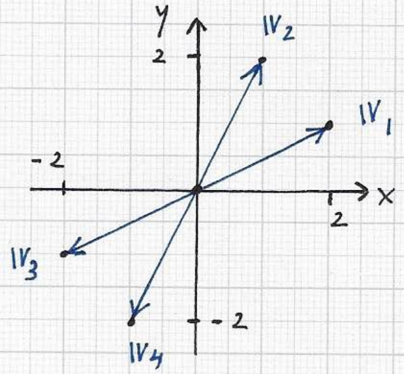
- Ellipse: $\left(1 : \frac{\sqrt{\lambda_1}}{2} \cdot x\right)^2 + \left(1 : \frac{\sqrt{\lambda_2}}{2} \cdot y\right)^2 = 1$

$$\Leftrightarrow (1:2 \cdot x)^2 + (1:1 \cdot y)^2 = 1$$

$$\Leftrightarrow \left(\frac{x}{2}\right)^2 + y^2 = 1$$

■ DIMENSION REDUCTION – Cont'd.

• Ex.: a slightly more complicated 2D example ...



– positional vectors after mean-subtraction:

$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $v_3 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$, $v_4 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$

– (covariance matrix) $C =$

$$\begin{pmatrix} 2 & 1 & -2 & -1 \\ 1 & 2 & -1 & -2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ -2 & -1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 10 & 8 \\ 8 & 10 \end{pmatrix}$$

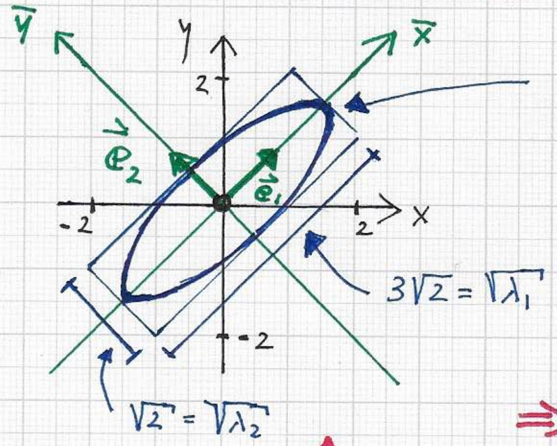
– characteristic polynomial $p(\lambda) =$

$$\begin{vmatrix} 10-\lambda & 8 \\ 8 & 10-\lambda \end{vmatrix} = (10-\lambda)^2 - 64$$

– C's eigenvalues: $(10-\lambda)^2 = 64 \Rightarrow \lambda_1 = 18$, $\lambda_2 = 2$

– C's eigenvectors: $\begin{pmatrix} 10 & 8 \\ 8 & 10 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \lambda_i \begin{pmatrix} x_i \\ y_i \end{pmatrix}$, $i = 1, 2$

$\Rightarrow \vec{e}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$, $\vec{e}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$ (normalized)



– Ellipse: directions and lengths of axes defined by $\lambda_1, \lambda_2, \vec{e}_1, \vec{e}_2$

The $\bar{x}$-coordinates are "more important" than the $\bar{y}$-coordinates.

$\Rightarrow$ REDUCE DATA TO $\bar{x}$-DIMENSION ONLY!

$\Rightarrow$ PCA defines a new data-inherent coordinate system. Axes and coordinates relative to this new system have decreasing degrees of "importance" – based on eigenvalues.