

# Segmentación

## Modelos Activos de Forma

### Lección 03.2

Dr. Pablo Alvarado Moya

MP6127 Visión por Computadora  
Programa de Maestría en Electrónica  
Énfasis en Procesamiento Digital de Señales  
Escuela de Ingeniería Electrónica  
Tecnológico de Costa Rica

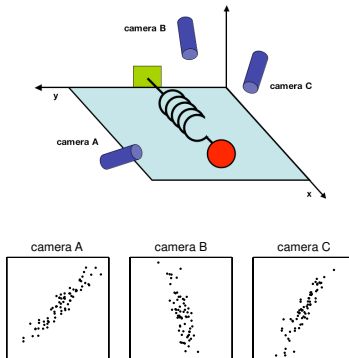
I Cuatrimestre 2013



- El Análisis de Componentes Principales (ACP, o PCA) es una herramienta estándar del análisis de datos.
- Aplicaciones desde neurociencias hasta gráficos por computador.
- Objetivo: reducir dimensiones de datos
- Se revelan relaciones ocultas en esos datos.
- Referencia:  
Shlens, J. A Tutorial on Principal Component Analysis.

- Supóngase experimento en donde se miden varias cantidades (e. g. espectros, tensiones eléctricas, velocidades, etc.)
- Los datos medidos suelen ser difusos, redundantes, poco claros
- Rescatar interrelaciones entre mediciones es un obstáculo en ciencias con validaciones empíricas.

- Supóngase el montaje de un resorte ideal.



Shlens

Bola de masa  $m$  sujeta a resorte sin masa ni fricción. La bola se libera a pequeña distancia del equilibrio. La bola oscila entonces en el eje  $x$ .

- Se ignora cuáles ejes son relevantes para la medición.

- Se toman medidas espaciales con 3 cámaras.
- Cámaras en 3 posiciones y orientaciones **arbitrarias**.
- La pregunta es  
¿cómo obtener de todas las mediciones la ecuación que pone en evidencia la única dependencia de  $x$ ?
- Las mediciones reales están contaminadas con ruido, y sufren condiciones no ideales.

- Meta del ACP: identificar base con “mayor sentido” para reexpresar los datos.
  - Con nueva base debe ser fácil filtrar efectos del ruido
  - Nueva base debe expresar estructuras ocultas en los datos
- Por ejemplo:  
Descubrir de las mediciones que  $\underline{x}$  es la dimensión de interés.

- Cada dato del experimento se denomina **muestra**.
- Cada muestra es un vector con todas las mediciones (tensión eléctrica, posición, etc.)
- Captura por intervalo de tiempo produce conjunto de muestras
- En el ejemplo: cámara  $i$  reporta posición  $(x_i, y_i)$ , así:

$$\underline{\mathbf{x}} = [x_A \quad y_A \quad x_B \quad y_B \quad x_C \quad y_C]^T$$

- Cada muestra  $\underline{\mathbf{x}}$  es un vector en un espacio  $m$ -dimensional, con  $m$  el número de mediciones tomadas.
- Espacio vectorial engendrado por base ortogonal



- Por convención, base **canónica** formada por los vectores

$$\underline{\mathbf{b}}_1 = [1, 0, 0, \dots, 0]^T$$

$$\underline{\mathbf{b}}_2 = [0, 1, 0, \dots, 0]^T$$

$$\vdots$$

$$\underline{\mathbf{b}}_n = [0, 0, 0, \dots, 1]^T$$

- La matriz de proyección a esta base canónica es

$$\mathbf{B} = \begin{bmatrix} \underline{\mathbf{b}}_1^T \\ \underline{\mathbf{b}}_2^T \\ \vdots \\ \underline{\mathbf{b}}_n^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}$$

- La pregunta es ahora

¿Existe alguna combinación **lineal** de la base original que exprese “mejor” el conjunto de datos?

(ver `pca.m`)

- Sea  $\mathbf{X}$  el conjunto original de  $m$  datos, donde cada columna es una muestra.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1i} & \cdots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2i} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{ni} & \cdots & x_{nm} \end{bmatrix}$$

- En ejemplo de resorte  $\mathbf{X}$  es una matriz de  $6 \times 72000$ .
- Sea  $\mathbf{Y}$  datos proyectados con transformación lineal  $\mathbf{P}$ :

$$\mathbf{Y} = \mathbf{P}\mathbf{X}$$

- $\mathbf{P}$  es una matriz que transforma  $\mathbf{X}$  en  $\mathbf{Y}$

- Sea  $\underline{\mathbf{p}}_i^T$  una fila de  $\mathbf{P}$
- Sea  $\underline{\mathbf{x}}_j$  una columna de  $\mathbf{X}$  (la  $j$ -ésima muestra)
- Sea  $\underline{\mathbf{y}}_j$  una columna de  $\mathbf{Y}$  (la  $j$ -ésima muestra)
- Geométricamente,  $\mathbf{P}$  rota y estira cada dato.
- Filas de  $\mathbf{P}$ : nueva base del espacio vectorial

$$\mathbf{PX} = \begin{bmatrix} \underline{\mathbf{p}}_1^T \\ \vdots \\ \underline{\mathbf{p}}_m^T \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_1 & \cdots & \underline{\mathbf{x}}_n \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} \underline{\mathbf{p}}_1^T \underline{\mathbf{x}}_1 & \cdots & \underline{\mathbf{p}}_1^T \underline{\mathbf{x}}_n \\ \vdots & \ddots & \vdots \\ \underline{\mathbf{p}}_m^T \underline{\mathbf{x}}_1 & \cdots & \underline{\mathbf{p}}_m^T \underline{\mathbf{x}}_n \end{bmatrix}$$

- La  $i$ -ésima columna de  $\mathbf{Y}$

$$\underline{\mathbf{y}}_i = \begin{bmatrix} \underline{\mathbf{p}}_1^T \underline{\mathbf{x}}_i \\ \underline{\mathbf{p}}_2^T \underline{\mathbf{x}}_i \\ \vdots \\ \underline{\mathbf{p}}_m^T \underline{\mathbf{x}}_i \end{bmatrix}$$

se conforma por la proyección del  $i$ -ésimo vector original  $\underline{\mathbf{x}}_i$  sobre cada uno de los vectores  $\underline{\mathbf{p}}_j$

Las preguntas que quedan por responder son

- 1 ¿cómo deben ser elegidos los vectores de la nueva base  $\mathbf{P}$ ?
- 2 ¿qué es bueno rescatar de  $\mathbf{X}$ ?, o similar
- 3 ¿qué debería contener  $\mathbf{Y}$ ?

- Valor medio  $\underline{\mu}$  del conjunto de datos es:

$$\underline{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

- Varianza  $\sigma^2$  es promedio de desviaciones respecto a  $\underline{\mu}$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \underline{\mu}\|_2^2 = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \underline{\mu})^T (\mathbf{x}_i - \underline{\mu})$$

- El cálculo de estas magnitudes se puede realizar con **un** solo recorrido por el conjunto de datos.
- **¿Cómo?**

- Demostración:

$$\begin{aligned}
 \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (\underline{\mathbf{x}}_i - \underline{\boldsymbol{\mu}})^T (\underline{\mathbf{x}}_i - \underline{\boldsymbol{\mu}}) \\
 &= \frac{1}{n} \sum_{i=1}^n \left( \underline{\mathbf{x}}_i^T \underline{\mathbf{x}}_i - \underline{\mathbf{x}}_i^T \underline{\boldsymbol{\mu}} - \underline{\boldsymbol{\mu}}^T \underline{\mathbf{x}}_i + \underline{\boldsymbol{\mu}}^T \underline{\boldsymbol{\mu}} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left( \|\underline{\mathbf{x}}_i\|^2 - 2\underline{\boldsymbol{\mu}}^T \underline{\mathbf{x}}_i + \|\underline{\boldsymbol{\mu}}\|^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \|\underline{\mathbf{x}}_i\|^2 - 2\underline{\boldsymbol{\mu}}^T \frac{1}{n} \sum_{i=1}^n \underline{\mathbf{x}}_i + \|\underline{\boldsymbol{\mu}}\|^2 \\
 &= \frac{1}{n} \sum_{i=1}^n \|\underline{\mathbf{x}}_i\|^2 - \|\underline{\boldsymbol{\mu}}\|^2
 \end{aligned}$$



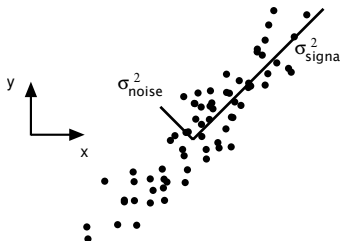
- Deben acumularse las magnitudes de los vectores y su suma

- Ruido de medición debe ser lo suficientemente bajo para poder rescatar información.
- Calidad de los datos (o señal) se mide con la tasa de señal a ruido (SNR), descrita a través de una tasa de varianzas:

$$SNR = \frac{\sigma_{\text{señal}}^2}{\sigma_{\text{ruido}}^2}$$

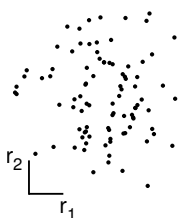
- $SNR \gg 1$  indica mediciones de alta precisión
- Baja SNR indica datos ruidosos.

- Volviendo al ejemplo del resorte.

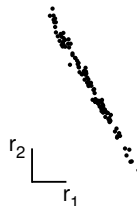


- Resorte mueve bola en línea recta  $\Rightarrow$  tres cámaras deberían observar un movimiento en línea recta.
- Toda desviación de la línea recta es ruido.
- Si señal es suficientemente fuerte, entonces dirección de **mayor varianza** coincide con la **señal**.
- Se busca rotación base canónica para alinearlos con los ejes de mayor varianza.

- Otro factor a considerar es la **redundancia**.
- En el ejemplo de las cámaras es notorio, pues cada cámara registra la misma información.
- Obsérvese que incluso en una misma imagen, si sube  $x$ , también sube  $y$ .
- Posibles dependencias entre dos variables  $r_1$  y  $r_2$ :



baja redundancia



alta redundancia

- En ejemplo, una dimensión es suficiente para describir los datos altamente redundantes (se puede reducir la dimensión)
- Dos dimensiones serían necesarias si datos en  $r_1$  y  $r_2$  no redundan.

- Considérense dos conjuntos de mediciones con media cero

$$A = \{a_1, a_2, \dots, a_n\} \quad B = \{b_1, b_2, \dots, b_n\}$$

- Las varianzas de los datos  $A$  y  $B$  por separado son

$$\sigma_A^2 = \frac{1}{n} \sum_i a_i^2, \quad \sigma_B^2 = \frac{1}{n} \sum_i b_i^2$$

- La **covarianza** entre  $A$  y  $B$  se generaliza como

$$\sigma_{AB}^2 = \frac{1}{n} \sum_i a_i b_i$$

- La covarianza mide el grado de dependencia lineal entre dos variables:
  - un valor positivo grande indica datos correlacionados positivamente
  - un valor negativo grande indica datos correlacionados negativamente
  - es cero si  $A$  y  $B$  no están correlacionados (no hay redundancia)
  - $\sigma_{AB}^2 = \sigma_A^2$  si  $A = B$
- Valor absoluto de covarianza indica grado de redundancia

- Expresando  $A$  y  $B$  como vectores:

$$A \rightarrow \underline{\mathbf{a}}^T = [a_1, a_2, \dots, a_n]$$

$$B \rightarrow \underline{\mathbf{b}}^T = [b_1, b_2, \dots, b_n]$$

se replantea la covarianza como

$$\sigma_{\underline{\mathbf{a}}\underline{\mathbf{b}}}^2 = \frac{1}{n} \underline{\mathbf{a}}^T \underline{\mathbf{b}}$$



- Lo anterior se puede generalizar de dos a cualquier número de vectores  $\underline{\mathbf{x}}_j$ : Dada la matriz  $\mathbf{X}$  de dimensiones  $m \times n$

$$\mathbf{X} = \begin{bmatrix} \underline{\mathbf{x}}_1^T \\ \underline{\mathbf{x}}_2^T \\ \vdots \\ \underline{\mathbf{x}}_m^T \end{bmatrix}$$

- Las filas de  $\mathbf{X}$  corresponden a todas las mediciones de un tipo particular (por ejemplo, eje x de la cámara A).
- Cada columna de  $\mathbf{X}$  corresponde a un conjunto de mediciones de una única muestra.
- La matriz de covarianza  $\Sigma_{\mathbf{X}}$  se define entonces como

$$\Sigma_{\mathbf{X}} = \frac{1}{n} \mathbf{X} \mathbf{X}^T$$

- El elemento  $(i, j)$  de la matriz  $\Sigma_{\mathbf{x}}$  es el producto punto entre el  $i$ -ésimo tipo de medición y el  $j$ -ésimo tipo.
- Esta matriz es
  - Cuadrada ( $m \times m$ ) y simétrica
  - La diagonal está formada por las varianzas de cada tipo de medición
  - Los términos fuera de la diagonal son las covarianzas entre tipos de medición.
- Los valores de covarianzas reflejan niveles de ruido y redundancia en las mediciones.
- En la diagonal, valores de gran magnitud indican estructuras relevantes
- Fuera de la diagonal, valores grandes en magnitud indican alta redundancia.

# Diagonalización de la matriz de covarianza

- Es de interés minimizar la redundancia y maximizar la señal de los datos de salida  $\mathbf{Y}$ , y por lo tanto
  - Todos los datos fuera de la diagonal de  $\Sigma_{\mathbf{Y}}$  deben ser cero, por lo que  $\Sigma_{\mathbf{Y}}$  es diagonal
  - Cada dimensión sucesiva de  $\mathbf{Y}$  deberá estar ordenada de acuerdo a la varianza.
- La matriz  $\Sigma_{\mathbf{X}}$  debe ser entonces **diagonalizada**
- El ACP asume que los vectores de la nueva base son ortonormales, y éstos constituyen los **componentes principales**.
- La varianza en cada componente indica qué tan importante es el componente.

- Se ha supuesto que el problema es **lineal**  
Un cambio de base expresado como producto de matrices puede solucionar el problema.
- Varianzas grandes indican estructura importante.  
Esto supone que se tiene alto SNR, para que las varianzas grandes estén asociadas con la señal.
- Componentes principales son ortogonales.  
Esto permite encontrar una solución utilizando álgebra lineal.
- Lo anterior se cumple si la distribución de los datos es normal

- El problema que se plantea es:  
dado el conjunto de datos  $\mathbf{X}$  (matriz  $m \times n$ , cada columna representa una muestra de  $m$  dimensiones, se cuenta con  $n$  muestras), encuéntrase una matriz ortogonal  $\mathbf{P}$  que transforma los datos a  $\mathbf{Y} = \mathbf{P}\mathbf{X}$ , de tal modo que

$$\Sigma_{\mathbf{Y}} = \frac{1}{n} \mathbf{Y}\mathbf{Y}^T$$

sea una matriz diagonal.

- Las filas de  $\mathbf{P}$  son los componentes principales de  $\mathbf{X}$

- Se parte de  $\Sigma_Y$

$$\begin{aligned}\Sigma_Y &= \frac{1}{n} \mathbf{Y} \mathbf{Y}^T \\ &= \frac{1}{n} (\mathbf{P} \mathbf{X}) (\mathbf{P} \mathbf{X})^T \\ &= \frac{1}{n} \mathbf{P} \mathbf{X} \mathbf{X}^T \mathbf{P}^T \\ &= \mathbf{P} \left( \frac{1}{n} \mathbf{X} \mathbf{X}^T \right) \mathbf{P}^T \\ &= \mathbf{P} \Sigma_X \mathbf{P}^T\end{aligned}$$

que relaciona las matrices de covarianza de datos de entrada y salida

- Lo anterior se cumple si las columnas de  $\mathbf{P}^T$  son los **eigenvectores** de  $\mathbf{X}$ , y los elementos en la matriz diagonal  $\Sigma_Y$  los **eigenvalores**.

- En otras palabras, las filas de  $\mathbf{P}$  son los **eigenvectores**, que se ordenan de acuerdo a sus correspondientes **eigenvalores** de forma descendiente.
- Los componentes principales de  $\mathbf{X}$  están dados entonces por los eigenvectores de su matriz de covarianza  $\Sigma_{\mathbf{X}} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$
- El  $i$ -ésimo valor de la diagonal de  $\Sigma_{\mathbf{Y}}$  es la varianza de  $\mathbf{X}$  a lo largo del componente principal  $\mathbf{p}_i$ .

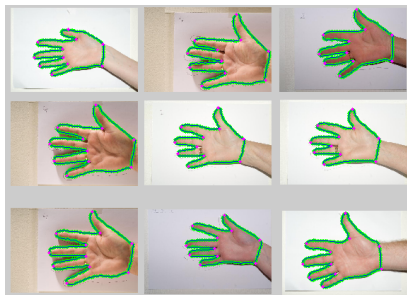
- 1 Substraer la media de todos los datos
- 2 Calcular la matriz de covarianza  $\Sigma_{\mathbf{X}}$
- 3 Calcular los eigenvectores y eigenvalores de  $\Sigma_{\mathbf{X}}$
- 4 Construir la matriz de proyección  $\mathbf{P}$  compuesta por el número de componentes principales deseados (seleccionados de acuerdo a los eigenvalores)
- 5 Proyectar los datos con  $\mathbf{Y} = \mathbf{P}\mathbf{X}$



- AAM <http://www.youtube.com/watch?v=I3YsqHCQB4k>
- Eigenfaces  
[http://www.youtube.com/watch?v=5nBL\\_u4MF0k](http://www.youtube.com/watch?v=5nBL_u4MF0k)

# Modelos Activos de Forma

- Propuesta de Cootes, Taylor, Cooper y Graham, 1995 (idea publicada en 1992)
- Idea: representar cada forma como un vector



- $s = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]^T$

# Modelo de Distribución de Puntos

- Modelo de distribución de puntos: ACP sobre todas formas de entrenamiento
- Se reducen dimensiones de representación
- Se supone que distribución de formas es normal
- Modelo extrae información de formas en datos de entrenamiento
- Es necesario alinear formas antes del ACP

- Alineación “normaliza” respecto a rotación, traslación y escalado
- Se usa el análisis de Procusto.

- Dadas dos formas  $\underline{\mathbf{x}}_1$  y  $\underline{\mathbf{x}}_2$ , se busca mapeo de rotación  $\theta$ , escala  $s$  y traslación  $\underline{\mathbf{t}} = [t_x, t_y]^T$  que mapee

$$\underline{\mathbf{x}}_2 \rightarrow M(\theta, s)[\underline{\mathbf{x}}_2] + \underline{\mathbf{t}}$$

con el operador

$$M(\theta, s) \begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix} = \begin{bmatrix} s \cos \theta & -s \sin \theta \\ s \sin \theta & s \cos \theta \end{bmatrix} \begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix}$$

y

$$\underline{\mathbf{t}} = [t_x, t_y, t_x, t_y, \dots, t_x, t_y]^T$$

que minimice

$$E = (\underline{\mathbf{x}}_1 - M(\theta, s)[\underline{\mathbf{x}}_2] - \underline{\mathbf{t}})^T \mathbf{W} (\underline{\mathbf{x}}_1 - M(\theta, s)[\underline{\mathbf{x}}_2] - \underline{\mathbf{t}})$$

con  $\mathbf{W}$  una matriz de ponderación de cada punto

- Por mínimos cuadrados se puede resolver este problema derivando  $E$  para cada parámetro  $\theta$ ,  $s$ ,  $t_x$  y  $t_y$  e igualando a cero.
- Lo anterior se resume en solucionar el sistema lineal:

$$\begin{bmatrix} X_2 & -Y_2 & W & 0 \\ Y_2 & X_2 & 0 & W \\ Z & 0 & X_2 & Y_2 \\ 0 & Z & -Y_2 & X_2 \end{bmatrix} \begin{bmatrix} a_x \\ a_y \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} X_1 \\ Y_1 \\ C_1 \\ C_2 \end{bmatrix}$$

donde  $a_x = s \cos \theta$ ,  $a_y = s \sin \theta$  y

$$X_i = \sum_{k=1}^n w_k x_{ik}$$

$$Y_i = \sum_{k=1}^n w_k y_{ik}$$

$$Z = \sum_{k=1}^n w_k (x_{2k}^2 + y_{2k}^2)$$

$$W = \sum_{k=1}^n w_k$$

$$C_1 = \sum_{k=1}^n w_k (x_{1k}x_{2k} + y_{1k}y_{2k})$$

$$C_2 = \sum_{k=1}^n w_k (y_{1k}x_{2k} - x_{1k}y_{2k})$$

# Alineación de todo el conjunto

- Los pesos **W** se eligen para dar más significancia a los puntos estables
- Algoritmo de alineación:
  - Alinee todas las formas con primera del conjunto
  - Repita
    - Calcule forma promedio de formas alineadas
    - Normalice orientación, escala y origen de la media actual
    - Reliniee todas las formas a la media actual
  - Hasta que proceso converja



- Luego del ACP sobre las formas alineadas se puede reconstruir cualquier forma a partir de los coeficientes de los componentes principales  $\underline{\mathbf{b}}$

$$\underline{\mathbf{x}} = \underline{\bar{\mathbf{x}}} + \mathbf{P}\underline{\mathbf{b}}$$

- $\mathbf{P}$  es la matriz de los primeros  $t$  eigenvectores
- $\underline{\mathbf{b}} = [b_1, b_2, \dots, b_t]^T$  es la parametrización de la forma
- Se restringen formas válidas logrando que

$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}$$

- $\lambda_k$  es el eigenvalor correspondiente al  $k$ -ésimo eigenvector

# Ajuste a imagen

- Una vez que se tienen los modelos entrenados, deben ajustarse a la imagen
- Se coloca una forma en algún lugar probable sobre la imagen
- Las marcas o hitos se ajustan buscando bordes (entre otras ideas)
- Una vez ajustados se determinan los parámetros **b** correspondientes
- Esos parámetros se restringen al rango válido
$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}$$
- La forma resultante se vuelve a ajustar
- En todo paso se deben determinar rotación, escala y traslación



- 1 Revisar tutorial sobre ACP de Shlens
- 2 Revisar artículo de Cootes et al. sobre ASM
- 3 Cree un modelo de forma para el tablero de ajedrez.

*Este documento ha sido elaborado con software libre incluyendo L<sup>A</sup>T<sub>E</sub>X, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, L<sup>T</sup>I-Lib-2, GNU-Make, Kazam, Xournal y Subversion en GNU/Linux*



Este trabajo se encuentra bajo una Licencia Creative Commons Atribución-NoComercial-LicenciarIgual 3.0 Unported. Para ver una copia de esta Licencia, visite <http://creativecommons.org/licenses/by-nc-sa/3.0/> o envíe una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

© 2013

Pablo Alvarado-Moya

Escuela de Ingeniería Electrónica

Instituto Tecnológico de Costa Rica