

Introducción al reconocimiento de patrones: Trabajo práctico 3, Análisis de Componentes Principales para reconocimiento de rostros

M. Sc. Saúl Calderón Ramírez
Instituto Tecnológico de Costa Rica,
Escuela de Computación, bachillerato en Ingeniería en Computación,
PAttern Recongition and MACHine Learning Group (PARMA-Group)

12 de septiembre de 2018

Fecha de entrega: 20 de Setiembre.

Entrega: Un archivo .zip con el código fuente LaTeX o Lyx, el pdf, y un script en MATLAB, debidamente documentado, con una función definida por ejercicio. A través del TEC-digital.

Modo de trabajo: Grupos de 3 personas.

1. Análisis de componentes principales para rostros

El reconocimiento de caras propuesto en [1] se basa en el análisis de componentes principales. Es entonces útil repasar en que consiste esta técnica muy utilizada en el reconocimiento de patrones y en general en el análisis de datos.

El ACP se basa en el cálculo de los auto-valores y auto-vectores. Los auto-vectores \vec{v} y auto valores λ cumplen la siguiente ecuación:

$$A\vec{v} = \lambda\vec{v} \rightarrow (A - \lambda I)\vec{v} = 0$$

Esta ecuación examina el comportamiento de la matriz A como una transformación de escalamiento (pues $\lambda \in \mathbb{R}$ es un escalar). Por ejemplo, en caso de que en la solución de tal sistema de ecuaciones con valores de λ muy altos se presenta un caso de escalamiento muy alto para el auto vector \vec{v} . Para encontrar los auto-valores λ , nos basamos en la suposición de que el sistema de ecuaciones anterior tiene infinitas soluciones, por lo que entonces el determinante de la ecuacion anterior debe ser cero, es decir:

$$\det(A - \lambda I) = 0$$

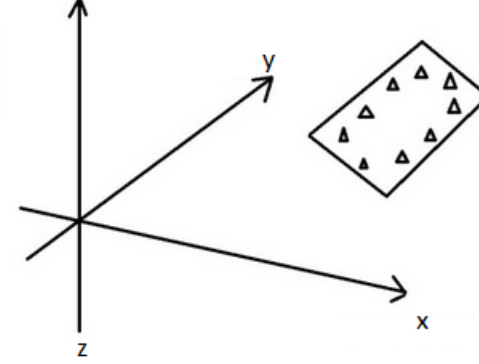


Figura 1: Datos de dimensionalidad 3.

La ecuación anterior denota el polinomio característico, de grado n , por lo que los autovalores de A corresponden a las raíces de su polinomio característico. Cada auto-valor λ_i es asociado a un auto-vector \vec{v}_i que se obtiene al resolver la ecuación 1 con el valor encontrado de λ_i .

En datos de dimensionalidades muy altas es frecuentemente útil encontrar cuales componentes (ejes o dimensiones de los datos) son los mas significativos, es decir, los principales. Por ejemplo en la Figura 1 se muestra gráficamente un conjunto de datos específico en tres dimensiones.

El ACP se basa en el concepto de encontrar los vectores o dimensiones que aportan mayor información en la distribución de los datos observados. Esto significa encontrar los «ejes» que impliquen una mayor variación de los datos respecto a cada uno de ellos, tal y como lo ilustra la Figura 2.

Los «ejes» con varianza baja respecto a los datos implican un aporte pequeño de información por lo que pueden ser descartados, como por ejemplo para el nuevo eje \vec{ev}_3 diagramado en la Figura 3.

Para encontrar tales «ejes», el ACP propone la descomposición en autovalores y auto-vectores de la matriz de covarianza Σ asociada a la matriz de datos A . La matriz de covarianza Σ para datos de dimensionalidad n es de dimensiones $n \times n$ con la diagonal especificando la variación de los datos en cada dimensión, y en entradas distintas a la diagonal definiendo la variación entre el par de dimensiones específicas.

Los «ejes» ejemplificados al corresponder a los auto-vectores de la matriz de covarianza tienen asociados auto-valores, los cuales definen el aporte en la distribución de los datos de su auto-vector. De esta manera es posible reducir

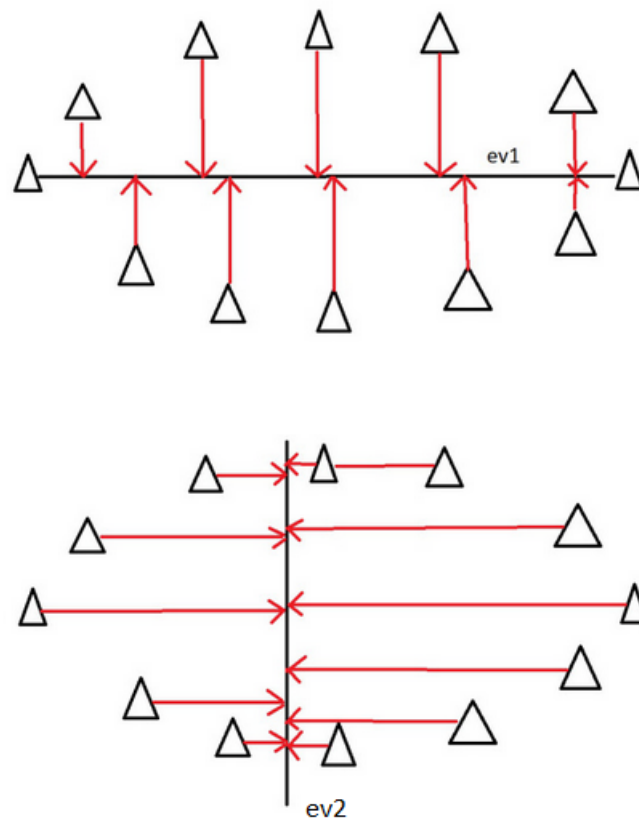


Figura 2: Ejes con mayor variación.

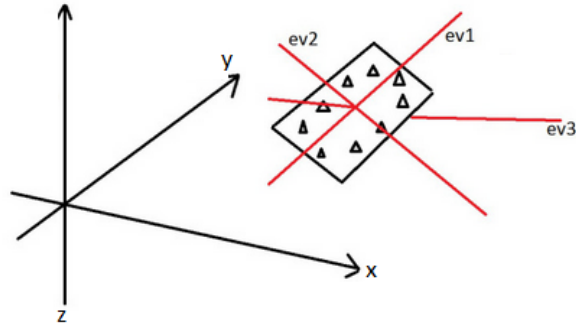


Figura 3: Autovectores de datos de dimensionalidad 3.

la dimensionalidad de los datos, creando un nuevo espacio vectorial el cual incluye los auto-vectores con los valores mayores de sus auto-valores. En el ejemplo explicado a lo largo de esta sub-sección tal reducción consiste en tomar los primeros 2 auto-vectores y se grafica en la Figura 4.

1.1. Auto-Caras para la clasificación

Lo propuesto en [1] se basa en el concepto de encontrar un nuevo espacio vectorial, con dimensionalidad reducida y mayor significancia en sus ejes, por lo que se descartan auto-vectores que aportan poca información (asociada al ruido). Utilizando los datos de entrenamiento, en este caso, correspondiente

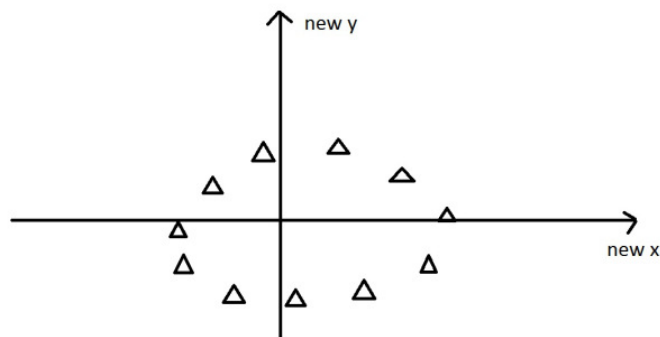


Figura 4: Nuevo espacio vectorial tomando los auto-vectores mas importantes.

a la base de datos con imagenes de rostros. Para la etapa de clasificación se utiliza tal espacio vectorial nuevo, se proyectan los nuevos datos y se calcula su distancia euclidiana y a partir de umbrales previamente establecidos clasificar si los nuevos datos pertenecen a una cara existente en la base de datos.

Más precisamente, la técnica de auto-caras para la clasificación de la pertenencia de una imagen digital de un rostro a un sujeto (lo cual define) implementa las siguientes etapas:

1. **Análisis de componentes principales:** El análisis de los componentes principales de un de $m = s \cdot p$ conjunto de datos, (con s clases y p muestras por clase) representado matricialmente como

$$X = \begin{bmatrix} | & | & \cdots & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \\ | & | & \cdots & | \end{bmatrix},$$

con $\vec{x}_i \in \mathbb{R}^n$ implica la creación de un nuevo espacio vectorial con dimensionalidad reducida $\vec{x}_i \in \mathbb{R}^{n'}$, donde las dimensiones que aportan menor cantidad de información son descartadas.

- a) Cálculo de la matriz de covarianza Σ para los datos en X .
 - b) Generación del nuevo espacio vectorial: Se calculan los auto-vectores $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_q\}$ y auto-valores correspondientes $\{\lambda_1, \lambda_2, \dots, \lambda_q\}$ y se eligen $n' < q$ auto-vectores más significativos para crear el nuevo espacio vectorial en $\mathbb{R}^{n'}$ espacioGenerado $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{n'}\}$ generado por tales auto-vectores.
 - c) Proyección de los datos en el espacio \mathbb{R}^n al nuevo espacio espacioGenerado $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{n'}\}$, con dimensionalidad $\mathbb{R}^{n'}$. Dependiendo del enfoque de clasificación supervisada no paramétrica a utilizar en la siguiente etapa, la proyección puede incluir todos los datos en la matriz X , o las muestras promedio $\vec{\mu}_1, \dots, \vec{\mu}_s$ de cada una de las s clases.
2. **Clasificación no paramétrica en el espacio vectorial de dimensionalidad reducida $\vec{x}_i \in \mathbb{R}^{n'}$:** La clasificación no paramétrica de una nueva muestra \vec{u} puede realizarse utilizando los siguientes enfoques:
 - a) Distancia al centroide de la clase μ_i , con $1 \leq i \leq s$ más cercana, usando una distancia euclidiana o $\ell = 2$.
 - b) K-vecinos más cercanos, calculando los k -vecinos más cercanos a la muestra \vec{u} y seleccionar la clase i con más muestras en tal conjunto.

1.1.1. Cálculo de la matriz de covarianza

Formalmente, se define cada imagen de las m caras o vectores muestra de entrenamiento $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m$, de s sujetos distintos y con p muestras por cada uno, de modo que $m = s \cdot p$. Cada imagen de rostro está compuesta por $n = a \times b$

pixeles en total, por lo que se supone que cada muestra es vectorizada, de modo que $\vec{x}_i \in \mathbb{R}^n$.

Para obtener la matriz de covarianza, se calcula primero la «cara promedio» o en general, «punto promedio», de la siguiente manera:

$$\vec{\mu} = \frac{1}{m} \sum_{i=1}^m \vec{x}_i$$

La matriz de diferencias respecto a la cara promedio, se define como

$$D = \begin{bmatrix} | & | & \dots & | \\ \vec{d}_1 & \vec{d}_2 & \dots & \vec{d}_m \\ | & | & \dots & | \end{bmatrix}$$

con $\vec{d}_i = \vec{x}_i - \vec{\mu}$, con cada diferencia $\vec{d}_i \in \mathbb{R}^n$ compuesta por n pixeles, por lo que entonces la matriz se define en un espacio $D \in \mathbb{R}^{n \times m}$. Dada la definición en términos vectoriales de la matriz de covarianza, **sin normalizar por $\frac{1}{m-1}$** pues lo que interesa son las direcciones de los autovectores que obtendremos posteriormente, y el coeficiente de normalización sólo cambia la magnitud de los autovalores, lo que nos ayuda a simplificar la notación:

$$\Sigma = \sum_{i=1}^m (\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T = \sum_{i=1}^m \vec{d}_i \vec{d}_i^T = D D^T$$

La matriz D es utilizada para construir la matriz de covarianza como $\Sigma_w = D D^T$. Sin embargo, de esta manera $\Sigma_w \in \mathbb{R}^{n \times n}$, lo cual hace el problema de encontrar los **auto-valores y auto-vectores computacionalmente costoso** para imágenes de resoluciones típicas (por ejemplo para una imagen de 640×480 pixeles, $n = 307200$). Veremos en la próxima subsección como de forma alternativa se puede utilizar la matriz de covarianza de los sujetos $\Sigma_v \in \mathbb{R}^{m \times m}$ calculando la matriz de covarianza como

$$\Sigma_v = D^T D$$

con una dimensionalidad menor usualmente, pues $m \ll n$.

El código de MATLAB muestra las distintas alternativas de calcular la matriz de covarianza:

```
%matriz de muestras m = 4
% muestras n = 3;
X = [3 2 5 7; 2 1 5 1; -5 -7 -8 -9];
%vector columna mu = mean(X')';
%matriz de diferencias D
D = X - repmat(mu, 1, m);
%Forma vectorial de la matriz de covarianza
%Para las cuatro muestras
Sigma1 = zeros(n, n)
```

```

for i = 1:m
    Sigma1 = Sigma1 + (D(:, i) * (D(:, i)')));
end
Sigma1 = (1 / (m - 1)) * Sigma1
%Forma matricial
Sigma2 = (1 / (m - 1)) * D * (D')
%matlab
Sigma3 = cov(X')

```

1.1.2. Generación del nuevo espacio vectorial

De esta manera, tal como se detalló en la Sub-sección anterior, los auto-valores λ_q y auto-vectores \vec{v}_q son calculados a partir de la matriz Σ_v , para un menor costo computacional. Cada auto-vector $\vec{v}_i \in \mathbb{R}^m$ satisface entonces la ecuación:

$$\Sigma_v \vec{v}_i = \lambda_i \vec{v}_i \Rightarrow D^T D \vec{v}_i = \lambda_i \vec{v}_i, \quad (1)$$

para los n **auto-vectores respectivos**, la anterior ecuación se puede reescribir matricialmente con

$$V = \begin{bmatrix} | & | & \dots & | \\ \vec{v}_1 & \vec{v}_2 & & \vec{v}_n \\ | & | & & | \end{bmatrix} \quad A = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

de la siguiente forma:

$$\Sigma_v V = AV. \quad (2)$$

Sin embargo nos sigue interesando el cálculo de los auto-vectores para la matriz de covarianza Σ_w , pues **estos describen la covarianza en cada pixel de las imágenes**. Para hacerlo, se sigue la ecuación 1 la cual multiplicando a ambos lados por D se obtiene que:

$$D D^T D \vec{v}_i = \lambda_i D \vec{v}_i, \quad (3)$$

con lo cual se puede observar que al reemplazar $\vec{w}_i = D \vec{v}_i \Rightarrow W = D V$ se obtiene que:

$$D D^T \vec{w}_i = \lambda_i \vec{w}_i, \quad (4)$$

lo que significa que el vector $\vec{w}_i \in \mathbb{R}^n$ (la cantidad de pixeles de la imagen de entrada) corresponde al auto-vector de la matriz de covarianza $\Sigma_w = D D^T$, para la cual deseamos conocer sus dimensiones principales, lo que hace que la ecuación anterior se reescriba como:

$$\Sigma_w \vec{w}_i = \lambda_i \vec{w}_i, \quad (5)$$

La matriz $W = [\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n]$ es entonces el conjunto de **auto-vectores** o **auto-caras** de la matriz Σ_w , para la cual nos evitamos calcular tales auto-vectores

directamente, ahorrando mucho del costo computacional. La dimensionalidad del nuevo espacio vectorial puede ser reducida incluso más utilizando los n' auto-vectores con auto-valores mayores, por lo que n' es un **parámetro de importancia en la construcción del nuevo espacio vectorial**, con lo que redefinimos a $W = [\vec{w}_1, \vec{w}_2, \dots, \vec{w}_{n'}]$, el cual genera un sub-espacio $E = \text{espacioGenerado} \{\vec{w}_1, \dots, \vec{w}_{n'}\}$. Observe que las auto-caras se calculan para todos los sujetos, y son independientes de la cantidad m total de muestras.

Para la etapa siguiente de clasificación (en la que se puede usar un enfoque de k vecinos más cercanos por ejemplo), será necesaria la proyección de los datos de entrenamiento con en el nuevo sub-espacio conformado por las auto-caras $W = [\vec{w}_1, \vec{w}_2, \dots, \vec{w}_{n'}]$. Para entender la proyección de un vector \vec{d}_i (correspondiente por ejemplo a una muestra de rostro en la base de entrenamiento, *normalizada* respecto al centroide de los datos al hacer $\vec{d}_i = \vec{x}_i - \mu$) a un sub-espacio vectorial $E = \text{espacioGenerado} \{\vec{w}_1, \dots, \vec{w}_{n'}\}$, recordemos que nos interesa calcular la magnitud de la proyección en cada eje del nuevo espacio

para cada muestra \vec{d}_i , para formar el vector $\vec{d}'_i = \begin{bmatrix} d'_{i,1} \\ \vdots \\ d'_{i,n'} \end{bmatrix}$ correspondiente a la auto-cara del sujeto i :

$$\begin{aligned} d'_{i,1} &= \vec{w}_1 \cdot \vec{d}_i = \vec{w}_1^T \vec{d}_i \\ &\vdots \\ d'_{i,n'} &= \vec{w}_{n'} \cdot \vec{d}_i = \vec{w}_{n'}^T \vec{d}_i \end{aligned}$$

Lo cual matricialmente, para todos los sujetos normalizados en D , se escribe como:

$$D' = W^T D$$

∴

El resultado de la etapa de entrenamiento necesario para la posterior clasificación de una nueva imagen de cara \vec{u} es la matriz de caras proyectadas D' , y la matriz de auto-caras W .

1.2. Clasificación

Para la etapa de clasificación es posible utilizar dos enfoques supervisados no paramétricos: K-vecinos más cercanos y la muestra promedio.

El algoritmo de K-vecinos más cercanos consiste en calcular las K muestras más cercanas a la nueva imagen a clasificar \vec{u} , y de tales muestras, contabilizar la clase C_i con mayor número de muestras, y tomar la decisión de la etiqueta a estimar $t = i$.

	$k = 2$	$k = 5$	$k = 7$	$k = 10$
10 %				
30 %				
50 %				
70 %				
90 %				

Cuadro 1: Experimentación para el clasificar de k vecinos más cercanos.

2. Experimentación

Realice los siguientes experimentos, donde por cada corrida, construya un conjunto de muestras de entrenamiento X_t y un conjunto de muestras de validación X_v de forma aleatoria, con 80 % de las muestras para entrenamiento y el resto para validación.

1. Para el clasificador de k vecinos más cercanos: construya una tabla donde cuantifique el porcentaje de muestras incorrectamente clasificadas variando la cantidad de vecinos y el porcentaje de auto-vectores a conservar, como se observa en la Tabla 1. Escoga la cantidad de vecinos más cercanos con mejores más resultados, y grafique el porcentaje de muestras incorrectamente clasificadas respecto al porcentaje de auto-vectores conservados.
2. Para el clasificador centroide por clase $\vec{\mu}_i$, con $1 \leq i \leq s$ más cercana, usando una distancia euclidiana o $\ell = 2$, grafique el porcentaje de muestras incorrectamente clasificadas respecto a ejecutar el clasificador con 10 %, 30 %, 50 %, 70 % y 90 % de los autovectores conservados.
3. Compare los resultados y defina el mejor clasificador.
4. Para el mejor clasificador realice y documente al menos 5 pruebas manuales con muestras aleatoriamente escogidas del conjunto de validación.

Referencias

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.