

# Aseguramiento de la Calidad del Software: Análisis de Componentes Principales

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

2 de julio de 2019

El siguiente documento presenta los aspectos teóricos del análisis de componentes principales, a implementarse en el proyecto semestral del curso

## 1. Aspectos básicos del álgebra lineal

Para los aspectos básicos de algebra lineal, necesarios para comprender la técnica de Análisis de Componentes Principales, se recomienda ver el apéndice de este documento.

### 1.1. Vectores

Tal como se mencionó, un vector de dimensionalidad  $n$  o con  $n$  componentes se define de la siguiente manera:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

donde se dice que el vector está definido en un espacio  $\mathbb{R}^n$ . Presenta un punto de origen  $A = (a_1, a_2, \dots, a_n)$  y un punto de destino o final  $B = (b_1, b_2, \dots, b_n)$  y viene entonces dado por:

$$\vec{v} = \overrightarrow{AB} = (b_1, b_2, \dots, b_n) - (a_1, a_2, \dots, a_n) = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \\ \vdots \\ b_n - a_n \end{bmatrix}$$

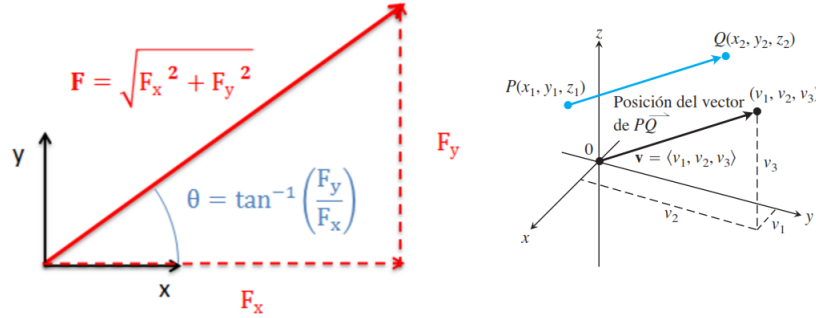


Figura 1: Vector con magnitud y dirección en  $\mathbb{R}^2$  y  $\mathbb{R}^3$ .

**Ilustración de conceptos con vectores en  $\mathbb{R}^2$**  Un vector tiene una dirección y una magnitud asociados, como lo sugiere el siguiente diagrama para un vector  $\vec{v} \in \mathbb{R}^2$ :

El **ángulo** por ejemplo de un vector  $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$  en un espacio  $\mathbb{R}^2$ , respecto al eje  $x$ , está dado por:

$$\theta = \arctan\left(\frac{v_2}{v_1}\right)$$

La **magnitud** se define, para un vector  $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$  en un espacio  $\mathbb{R}^2$  como :

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}$$

y en general para un vector en un  $\mathbb{R}^n$  como:

$$\|\vec{v}\| = \sqrt{v_1^2 + \dots + v_n^2}$$

recordemos además, que es un vector unitario todo aquel vector  $\hat{v}$  que cumpla con  $\|\hat{v}\| = 1$ .

**Producto punto o producto interno o producto escalar para un vector:** la función producto punto, para dos vectores  $\vec{w}$  y  $\vec{v}$  de dimensión  $n$  está dada por:

$$s = \vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} = v_1 w_1 + v_2 w_2 + \dots + v_n w_n = \sum_{i=1}^n v_i w_i$$

donde se dice que  $s$  es un escalar que pues  $s \in \mathbb{R}^1$ .

En el espacio euclidiano, el **producto punto** tiene la siguiente equivalencia geométrica:

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\| \cos(\theta)$$

donde el ángulo entre los vectores  $\vec{v}$  y  $\vec{w}$  está dado por  $\theta$ . El producto punto, gráficamente se refiere a la noción de la *sombra* o magnitud de la proyección del vector  $\vec{v}$  en  $\vec{w}$ , como muestra la Figura 2.

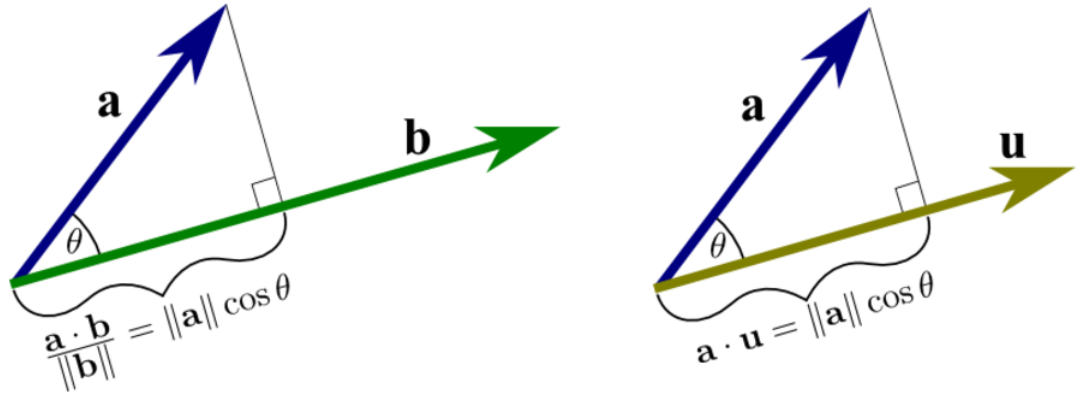


Figura 2: Magnitud de la proyección de los vectores  $\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos(\theta)$ , y los vectores  $\vec{a} \cdot \vec{u} = \|\vec{a}\| \cos(\theta)$ , con  $\|\vec{u}\| = 1$ , tomado de [http://mathinsight.org/dot\\_product](http://mathinsight.org/dot_product).

Esto quiere decir que si los dos vectores son **co-direccionales**  $\theta = 0 \Rightarrow \cos(\theta) = 1$  por lo que entonces:

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\|,$$

lo cual significa que si calculamos el producto punto del vector  $\vec{v}$  consigo mismo:

$$\vec{v} \cdot \vec{v} = \|\vec{v}\|^2,$$

por lo que podemos llegar a la definición de la **magnitud o norma** en  $\ell_2$  puede expresarse entonces en términos del producto punto como:

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}.$$

La **magnitud de un vector** puede interpretarse como la **proyección en el espacio  $\mathbb{R}^1$  en dirección del vector unitario  $\vec{v}_u$** , una operación básica que permite reducir la dimensionalidad de un vector (la reducción de la dimensionalidad es un concepto fundamental en el reconocimiento de patrones).

Por ejemplo, los vectores unitarios  $\hat{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ,  $\hat{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$  y  $\hat{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$  son vectores unitarios.

Si el ángulo entre los dos vectores **u ortogonales**  $\vec{v}$  y  $\vec{w}$  es de  $90^\circ$ , se tiene que entonces por la definición geométrica del producto punto:

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} = 0$$

Los vectores pueden dibujarse en MATLAB como sigue (en  $\mathbb{R}^2$ ):

```
M = [-0.4 0.7 0.2 ; -0.5 0.1 0.5];
plotv(M, ' -')
```

### 1.1.1. Independencia lineal y el rango de una matriz

Un conjunto de vectores  $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\} \subset \mathbb{R}^m$  se dice que es linealmente independiente, si ningún vector de tal conjunto puede ser representado como una combinación lineal del resto de vectores. De lo contrario, si uno de los vectores en tal conjunto puede ser representado como una combinación lineal del resto de vectores, entonces los vectores son **linealmente dependientes**, lo que se expresa como:

$$\vec{x}_j = \sum_{i=1}^{n-1} \alpha_i \vec{x}_i$$

para cualquier conjunto de valores escalares  $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$  se dice que el vector  $\vec{x}_j \in \mathbb{R}^m$  es linealmente dependiente de los vectores  $\vec{x}_i$ .

El **rango de columnas** de la matriz  $A \in \mathbb{R}^{m \times n}$  corresponde a la cantidad más grande de columnas en la matriz  $A$  linealmente independientes, de manera similar, el **rango de filas** se refiere a la cantidad más grande de filas en tal matriz linealmente independientes.

Para cualquier matriz  $A \in \mathbb{R}^{m \times n}$  se puede comprobar que el rango de filas y el de columnas es el mismo, por lo que entonces la cantidad de filas y columnas linealmente independiente se le refiere con el **rango**.

#### Ejemplo:

Observe la siguiente matriz:

$$\begin{bmatrix} 1 & 2 & -1 & 3 & -2 \\ 2 & 1 & 0 & 1 & 1 \\ 2 & 4 & -2 & 6 & -4 \\ 0 & 0 & 0 & 0 & 0 \\ 5 & 4 & -1 & 5 & 0 \end{bmatrix}$$

Fácilmente puede notarse que la fila  $f_3 = 2f_1$  y además que  $f_5 = 2f_2 + f_1$ , y que dado que la fila  $f_4$  es nula, entonces puede ser expresada en términos de cualquier otra fila en una combinación lineal. El rango por lo tanto de esta matriz es  $\text{rank}(A) = 2$ .

### 1.1.2. La matriz inversa

La inversa de la matriz cuadrada  $A \in \mathbb{R}^{n \times n}$  se denota como  $A^{-1}$  es la única matriz que cumple lo siguiente:

$$A^{-1}A = I = AA^{-1}$$

Nótese que no todas las matrices tienen inversas, por ejemplo las matrices no cuadradas no tienen inversas por definición, e incluso, pueden existir matrices cuadradas sin inversas.

- Se dice que  $A$  es una matriz **invertible** o no singular si  $A^{-1}$  existe, si la matriz  $A$  presenta **rango completo**.

- Si la matriz  $A^{-1}$  no existe, se dice que la matriz es **no invertible** o singular, y el **rango es incompleto**.

### Ejemplo 1:

Para un sistema de ecuaciones como el siguiente:

$$\begin{aligned} 2x_1 - x_2 &= -1 \\ 4x_1 - 2x_2 &= -2 \end{aligned}$$

Es posible expresarlo en términos matriciales como:

$$A\vec{x} = \vec{b} \Rightarrow \vec{x} = A^{-1}\vec{b}$$

con  $A \in \mathbb{R}^{2 \times 2}$ ,  $\vec{x}, \vec{b} \in \mathbb{R}^{2 \times 1}$  con  $A = \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$ ,  $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  y  $\vec{b} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$ . Si calculamos el rango de  $A$  haciendo:

$$\begin{aligned} A &= \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}; \\ \text{rank}(A) & \end{aligned}$$

resulta que  $\text{rank}(A) = 1$ , por lo que entonces la matriz es de rango incompleto, y por ende no invertible. Si graficamos las dos ecuaciones como rectas, veremos que en realidad, de forma normalizada, **se representa la misma línea**, lo que en términos de tales ecuaciones significa que la ecuación 1 es combinación lineal de la otra. El sistema tiene entonces infinitas soluciones al existir al menos una ecuación que esa combinación lineal de otras en el sistema, puesto que es suficiente con que:  $2x_1 = x_2 - 1$ . **Se puede demostrar entonces que en general, si el sistema de ecuaciones  $A \in \mathbb{R}^{n \times n}$  tiene infinitas soluciones, la matriz  $A$  es de rango incompleto.**

### 1.1.3. La matriz transpuesta

La transpuesta de una matriz es el resultado de cambiar las filas a columnas. Sea una matriz  $A \in \mathbb{R}^{m \times n}$ , su transpuesta se escribe como  $A^T \in \mathbb{R}^{n \times m}$  y sus entradas están dadas por:

$$(A^T)_{i,j} = A_{j,i}.$$

Las siguientes son propiedades de la transpuesta:

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A + B)^T = A^T + B^T$ .

### 1.1.4. Matrices simétricas

Una matriz cuadrada  $A \in \mathbb{R}^{n \times n}$  es simétrica si  $A = A^T$  y es anti simétrica si  $A = -A^T$ . Para toda matriz  $A \in \mathbb{R}^{n \times n}$  es fácil demostrar que la matriz  $A + A^T$

es simétrica y la matriz  $A - A^T$  es anti-simétrica, por lo que se puede seguir que cualquier matriz cuadrada puede expresarse en términos de una matriz simétrica y anti-simétrica:

$$A = \frac{1}{2} (A + A^T) - \frac{1}{2} (A - A^T).$$

Se define entonces el conjunto de matrices simétricas de dimensiones  $n \times n$  como  $\mathbb{S}^n$  por lo que  $A \in \mathbb{S}^n$  si es simétrica. Las matrices simétricas son muy frecuentes en el reconocimiento de patrones, y presentan una serie de propiedades muy útiles que veremos más adelante.

### 1.1.5. Rango y espacio nulo de la matriz

Un **espacio generado** de un conjunto de vectores base  $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$   $\vec{a}_i \in \mathbb{R}^n$  es el conjunto de vectores que pueden ser expresados como combinación lineal de tales vectores base:

$$\text{espacioGenerado}(\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}) = \left\{ \vec{v} : \vec{v} = \sum_{i=1}^m x_i \vec{a}_i \quad x_i \in \mathbb{R}^1 \right\}.$$

Puede demostrarse que si el conjunto de vectores  $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$   $\vec{a}_i \in \mathbb{R}^n$  es **linealmente independiente** (con  $m \geq n$ ), el espacio generado por tal conjunto de vectores es:

$$\text{espacioGenerado}(\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}) = \mathbb{R}^n.$$

Por ejemplo, los vectores unitarios anteriormente presentados  $\hat{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ,  $\hat{j} =$

$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$  y  $\hat{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$  son linealmente independientes, por lo que entonces es fácil observar que la combinación lineal de tales vectores puede generar cualquier vector en el espacio  $\mathbb{R}^3$ . Por ejemplo, un vector  $\vec{v} = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$  se puede representar como:

$$\vec{v} = 3\hat{i} + 5\hat{j} + 7\hat{k} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

por lo que entonces  $\vec{v} \in \text{espacioGenerado}(\{\vec{i}, \vec{j}, \vec{k}\}) = \mathbb{R}^3$ , con en este caso  $x_1 = 3$ ,  $x_2 = 5$  y  $x_3 = 7$ .

El **espacio de columnas** de una matriz  $A \in \mathbb{R}^{m \times n}$  denotado como  $\mathcal{C}(A)$  corresponde al espacio generado por las columnas de la matriz  $A$ , lo cual se representa como sigue:

$$\mathcal{C}(A) = \{ \vec{v} \in \mathbb{R}^m : \vec{v} = A \vec{x}, \vec{x} \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m} \},$$

donde recordemos que la multiplicación matricial  $A \vec{x}$  corresponde a una combinación lineal del vector  $\vec{x}$ :

$$A \vec{x} = \begin{bmatrix} | & | & \cdots & | \\ \vec{a}_{:,1} & \vec{a}_{:,2} & \cdots & \vec{a}_{:,n} \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 [\vec{a}_{:,1}] + x_2 [\vec{a}_{:,2}] + \cdots + x_n [\vec{a}_{:,n}],$$

por lo que entonces el espacio de columnas de la matriz  $A$  equivale a:

$$\mathcal{C}(A) = \text{espacioGenerado}(\{\vec{a}_{:,1}, \vec{a}_{:,2}, \dots, \vec{a}_{:,n}\}) = \left\{ v : v = \sum_{i=1}^n x_i \vec{a}_{:,i} \quad x_i \in \mathbb{R}^1 \right\}.$$

**La proyección de un vector**  $\vec{y} \in \mathbb{R}^n$  en el espacio generado por el conjunto de vectores  $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$   $\vec{a}_i \in \mathbb{R}^n$  corresponde al vector  $\vec{v} \in \text{espacioGenerado}(\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\})$  tal que  $\vec{v} \in \mathbb{R}^n$  esté lo más cerca posible del vector  $\vec{y} \in \mathbb{R}^n$ , medido con por ejemplo una norma euclidiana  $\|\vec{v} - \vec{y}\|_2$  y se puede definir formalmente como:

$$\text{proy}(\vec{y}; \{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}) = \underset{\vec{v} \in \text{espacioGenerado}(\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\})}{\text{argmin}} \|\vec{v} - \vec{y}\|_2.$$

Este tema se retomará al final del presente documento, una vez que se haya definido el concepto de gradiente matricial y se demostrará que:

$$\text{proy}(\vec{y}; A) = \underset{\vec{v} \in \mathcal{C}(A)}{\text{argmin}} \|\vec{v} - \vec{y}\|_2 = A (A^T A)^{-1} A^T \vec{y}$$

Para el caso en que  $A$  está formada por una sola columna  $\vec{a} \in \mathbb{R}^m$  (correspondiente a un espacio generador de un vector), se tiene el caso especial de la proyección de un vector sobre otro vector:

$$\text{proy}(\vec{y}; \vec{a}) = \frac{\vec{a} \vec{a}^T}{\vec{a}^T \vec{a}} \vec{y}$$

Observe que en tal caso de fijar un conjunto generador de un solo vector, el subespacio generado corresponde únicamente al escalamiento de tal vector, pero la dimensionalidad del vector proyectado tiene la misma dimensionalidad original (por lo que se denomina una proyección a un sub-espacio). La Figura 3 muestra la proyección de un vector  $\vec{a}$  sobre otro vector  $\vec{b}$ .

```
function proyectar
    v1 = [3; 7];
    v2 = [9; 1];
    proy = proyectarVector(v1, v2);
    figure;
    plotv([proy v1]);
    figure;
    plotv([v2 v1]);
```

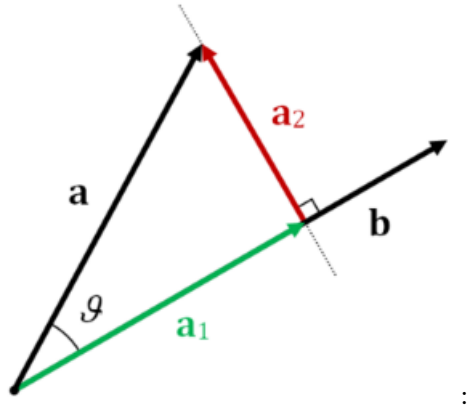


Figura 3: Proyección de vector  $\vec{a}$  sobre  $\vec{b}$ .

```

end
function proyec = proyectarVector(b, a)
    %proyecta b sobre a
    coefMatricial = ((a * a') / (a' * a));
    proyec = coefMatricial * b;
end

```

El **espacio nulo** de una matriz  $A \in \mathbb{R}^{m \times n}$ , se define como el conjunto de todos los vectores que al multiplicarse con la matriz  $A$  resultan en 0, y se denota como

$$\mathcal{N}(A) = \{\vec{x} \in \mathbb{R}^n : A\vec{x} = 0\}$$

**Ejemplo 1** ( $m = n$ , igual número de vectores en la base que dimensionalidad):

Sean los vectores  $\vec{a}_1 = \begin{bmatrix} 0,5 \\ 0 \\ 0 \end{bmatrix}$ ,  $\vec{a}_2 = \begin{bmatrix} 0 \\ 0,25 \\ 0 \end{bmatrix}$  y  $\vec{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$ , los cuales forman la matriz:

$$A = \begin{bmatrix} 0,5 & 0 & 0 \\ 0 & 0,25 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

determine el vector de proyección  $\text{proy}(\vec{y}; A) \in \mathbb{R}^3$ , para el caso en que  $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ . Observe que los vectores son linealmente independientes, y además, la cantidad de vectores  $m$  es menor a la dimensionalidad  $n$  de  $\vec{y}$ . La proyección de  $\vec{y}$  sobre el espacio de columnas de  $A$  está dado por el vector que resulta de:

$$\text{proy}(\vec{y}; A) = x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3$$



$$\Rightarrow \text{proy}(\vec{y}; A) = 2 \begin{bmatrix} 0,5 \\ 0 \\ 0 \end{bmatrix} + 8 \begin{bmatrix} 0 \\ 0,25 \\ 0 \end{bmatrix} + 1,5 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

por lo que entonces en este caso,  $x_1 = 2$ ,  $x_2 = 8$ ,  $x_3 = 1,5$  son los coeficientes que permiten calcular el vector proyección  $\text{proy}(\vec{y}; A)$  en el espacio generado por las columnas de  $A$ . Si usamos la fórmula para determinar tal vector proyección:

$$\text{proy}(\vec{y}; A) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \underset{\vec{v} \in \mathcal{C}(A)}{\text{argmin}} \|\vec{v} - \vec{y}\|_2 = A (A^T A)^{-1} A^T \vec{y}$$

implementando el siguiente código de MATLAB:

```
A = [0.5 0 0; 0 0.25 0; 0 0 2];
y = [1; 2; 3];
proyY_A = A*inv(A' * A) * A' * y;
x = inv(A) * proyY_A;
%Otra manera de obtener x1
x1 = (dot(y, A(:, 1)))/norm(A(:, 1));
%equivale a
coefMatricial = ((A(:, 1) * A(:, 1)') / (A(:, 1)' * A(:, 1)));
y1 = coefMatricial * y;
x1 = norm(y1, 2);
```

Observe que como los vectores son independientes entre si, es posible encontrar un vector proyección que hace que  $\underset{\vec{v} \in \mathcal{C}(A)}{\text{argmin}} \|\vec{v} - \vec{y}\|_2$ , por lo cual en este caso  $\|\text{proy}(\vec{y}; A) - \vec{y}\|_2 = 0$ .

De no conocer los coeficientes  $x_1 = 2$ ,  $x_2 = 8$ ,  $x_3 = 1,5$ , los mismos se pueden calcular siguiendo la ecuación:

$$\text{proy}(\vec{y}; A) = A \vec{x} \Rightarrow A^{-1} \text{proy}(\vec{y}; A) = \vec{x}$$

Lo cual para este caso resulta en  $\vec{x} = [2 \ 8 \ 1,5]^T$ . **La magnitud de la proyección en cada uno de los vectores de la base viene dada por:**

$$\|\text{proy}(\vec{y}; \vec{a}_1)\| = \left| \frac{\vec{y} \cdot \vec{a}_1}{\|\vec{a}_1\|} \right| = \left\| \frac{\vec{a}_1 \vec{a}_1^T}{\vec{a}_1^T \vec{a}_1} \vec{y} \right\| \neq x_i$$

**Ejemplo 4 ( $m < n$ , menos vectores en la base que dimensionalidad):**

Sean los vectores  $\vec{a}_1 = \begin{bmatrix} 5 \\ 7 \\ 21 \end{bmatrix}$  y  $\vec{a}_2 = \begin{bmatrix} 0 \\ 13 \\ 9 \end{bmatrix}$ , los cuales forman la matriz:

$A = \begin{bmatrix} 5 & 0 \\ 7 & 13 \\ 21 & 9 \end{bmatrix}$  determine el vector de proyección  $\text{proy}(\vec{y}; A) \in \mathbb{R}^3$ , para el caso

en que  $\vec{y} = \begin{bmatrix} 1,2 \\ 1,3 \\ 1,5 \end{bmatrix}$ .

Sabemos que el vector proyección viene dado entonces por  $\text{proy}(\vec{y}; A) = A (A^T A)^{-1} A^T \vec{y}$ . Para calcular lo anterior, se necesita que  $A$  sea una matriz cuadrada, por lo que formalmente no se puede calcular la inversa de tal matriz. Es por ello que recurrimos al cálculo de la pseudo-inversa, usando el método de Moore-Penrose. Implementando el siguiente código de MATLAB, se obtiene que  $\text{proy}(\vec{y}; A) = [1,1542 \quad 13,3663 \quad 14,9695]^T$ . Obsérvese que en el caso en que se disponen menos vectores en la base respecto a la dimensionalidad del vector a proyectar, por lo que la proyección tiene un error distinto de cero.

```
A = [5 0; 7 13; 21 9];
y = [1.2; 1.3; 1.5];
proyY_A = A * pinv(A' * A) * A' * y;
x = pinv(A) * proyY_A;
u1 = dot(y, A(:, 1)) / norm(A(:, 1));
u2 = dot(y, A(:, 2)) / norm(A(:, 2));
figure;
quiver3(0,0,0, A(1,1), A(2,1), A(3,1));
hold on;
quiver3(0,0,0, A(1,2), A(2,2), A(3,2));
hold on;
quiver3(0,0,0, y(1), y(2), y(3));
hold on;
quiver3(0,0,0, u1, u2, 5);
```

Para reducir la dimensionalidad del vector  $\vec{y}$  en una dimensión, en este caso, se construye un vector  $\vec{u}$  en  $\mathbb{R}^2$  cuyos componentes están definidos por la proyección :

$$u_1 = \frac{\vec{y} \cdot \vec{a}_1}{\|\vec{a}_1\|} = 2,0534$$

$$u_2 = \frac{\vec{y} \cdot \vec{a}_2}{\|\vec{a}_2\|} = 1,9227$$

Observe que las operaciones anteriores pueden resultar en un número negativo, por lo que preservan la dirección del vector  $\vec{u}$  en  $\mathbb{R}^2$ , a diferencia de usar  $u_i = \left\| \frac{\vec{a}_i \vec{a}_i^T}{\vec{a}_i^T \vec{a}_i} \vec{y} \right\|$ . El vector  $\vec{x}$  en  $\text{proy}(\vec{y}; A) = A \vec{x}$  nos indica los coeficientes en un espacio expresado en términos de los vectores base de  $A$ , pero como seguimos dibujando en un espacio  $\mathbb{R}^2$  cuya base son los vectores unitarios  $\hat{i}, \hat{j}$ , la reducción de dimensionalidad se hace usando la proyección de  $\vec{y}$  sobre cada vector base  $\vec{a}_i$ .

### 1.1.6. Determinante de una matriz

El determinante de una matriz cuadrada  $A \in \mathbb{R}^{n \times n}$  es una función denotada con  $\det(A) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ . Antes de detallar la fórmula que define al determinante, examinaremos la interpretación geométrica del determinante. Sea una matriz compuesta por múltiples filas:

$$A = \begin{bmatrix} - & \vec{a}_{1,:}^T & - \\ - & \vec{a}_{2,:}^T & - \\ & \vdots & \\ - & \vec{a}_{n,:}^T & - \end{bmatrix}$$

considere el conjunto de puntos  $S \subset \mathbb{R}^n$  formado al tomar todas las combinaciones lineales posibles de los vectores fila  $\vec{a}_{i,:}^T$ , donde los coeficientes de tal combinación lineal cumplen que  $0 \leq \alpha_i \leq 1, i = 1, \dots, n$ , lo cual formalmente se denota como:

$$S = \left\{ \vec{v} \in \mathbb{R}^n : \vec{v} = \sum_{i=1}^n \alpha_i \vec{a}_{i,:}, \quad 0 \leq \alpha_i \leq 1, i = 1, \dots, n \right\}$$

El valor absoluto del determinante de la matriz  $A$ ,  $|\det(A)|$ , corresponde a una medida del “volumen” de todo el conjunto  $S$ .

Por ejemplo, para la matriz  $A \in \mathbb{R}^{2 \times 2}$ :

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$$

cuyos vectores fila están dados por:

$$\vec{a}_{1,:} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \vec{a}_{2,:} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

se muestra en la Figura 4, sombreado, el conjunto de puntos  $S$ . Observe que el punto “extremo”  $\vec{a}_{1,:} + \vec{a}_{2,:} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ , viene dado cuando  $\alpha_1 = \alpha_2 = 1$ . El determinante para una matriz de  $2 \times 2$  se define como:

$$\det \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = a d - b c$$

y para cualquier matriz de  $n \times n$  dimensiones, el determinante se define recursivamente como:

$$\det(A) = A_{1,1} \det(A_{\setminus 1, \setminus 1}) - A_{1,2} \det(A_{\setminus 1, \setminus 2}) + \dots \pm A_{1,n} \det(A_{\setminus 1, \setminus n})$$

lo cual es equivalente también a escoger cualquier fila o columna a eliminar:

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} A_{i,j} |A_{\setminus i, \setminus j}| = \sum_{j=1}^n (-1)^{i+j} A_{i,j} |A_{\setminus i, \setminus j}|$$

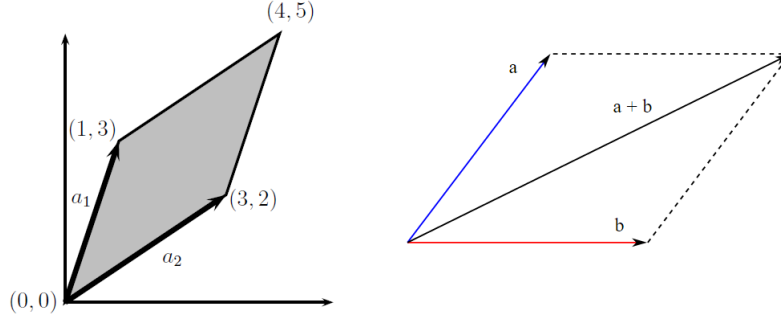


Figura 4: Región  $S$  de ejemplo.

Observe que el determinante consiste en la combinación lineal de los determinantes de las submatrices resultantes de eliminar la fila y columna  $i$  (denotado como  $\det(A_{\setminus i, \setminus j})$ ), multiplicado por el elemento  $A_{1,i}$ . Con la matriz de ejemplo  $A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$ , el determinante viene entonces dado por:  $\det(A) = 1 \cdot 2 - 3 \cdot 3 = -7$ , y tomando su valor absoluto, se tiene que  $|\det(A)| = 7$ , lo que corresponde al área del paralelogramo formado por el conjunto de puntos  $S$  (en  $n$  dimensiones, se refiere como paralelótopo).

Las siguientes son propiedades de la función determinante  $\det(A)$  para una matriz cuadrada  $A \in \mathbb{R}^{n \times n}$ :

- $\det(A) = 0$ , implica que  $A$  es una matriz singular (no invertible), por lo que entonces no tiene rango completo, y sus columnas son **linealmente dependientes**, lo que implica a demás la superficie  $S$  no tiene volumen, al ser los vectores combinación lineal. Ello también implica que el sistema de ecuaciones que representa  $A \in \mathbb{R}^{n \times n}$  tiene **infinitas soluciones**.

### 1.1.7. Autovalores y auto-vectores

Sea una matriz cuadrada  $A \in \mathbb{R}^{n \times n}$ , decimos que  $\lambda \in \mathbb{C}$  es un **auto-valor** o eigen-valor de  $A$  y el vector  $\vec{x} \in \mathbb{C}^n$  es su **auto-vector** o *eigen-vector* si:

$$A\vec{x} = \lambda\vec{x}, \quad \vec{x} \neq 0 \quad (1)$$

Intuitivamente, la ecuación anterior significa que la multiplicación de la matriz  $A$  por un vector  $\vec{x}$  es igual a la multiplicación de tal vector  $\vec{x}$  por el escalar  $\lambda$  también referido como el escalamiento del vector  $\vec{x}$ .

Los auto-vectores son vectores normalizados (con magnitud unitaria), puesto que cualquier vector escalado de  $\vec{x}$ ,  $\vec{v} = c\vec{x}$  hace que la ecuación  $A\vec{v} = \lambda\vec{v}$  se siga cumpliendo. Siguiendo la ecuación 1, se tiene que:

$$A\vec{x} - \lambda\vec{x} = (\lambda I - A)\vec{x} = 0, \quad \vec{x} \neq 0 \quad (2)$$

La ecuación anterior tiene solución no nula o no-cero si y solo si la matriz  $(\lambda I - A)$  tiene un espacio nulo no vacío, lo cual es el caso si y solo si tal matriz es singular (no-invertible), por lo que en términos del determinante debe cumplir que:

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

Esto pues en general, si un sistema de ecuaciones  $A\vec{x} = \vec{b}$ , se tiene que  $A$  es invertible, existe entonces una solución única  $\vec{x} = A^{-1}\vec{b}$ . **Si la matriz  $A$  no es invertible, existen múltiples e infinitas soluciones.**

De esta forma, con el cálculo del determinante, se construye el polinomio en términos de la variable  $\lambda$  y de grado  $n$ , para lo cual se encuentran las raíces de tal polinomio. Una vez conocidos los auto-valores  $\lambda$ , se procede a buscar sus autovectores correspondientes  $\vec{x}$ , resolviendo la ecuación matricial  $(\lambda I - A)\vec{x} = 0$ . Un ejemplo puede dilucidar mejor el procedimiento:

Sea la matriz

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

La matriz para cuyo espacio nulo se realizará el cálculo viene dada por:

$$(\lambda I - A) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} \lambda & -1 \\ 2 & \lambda + 3 \end{bmatrix}$$

y su determinante está dado entonces por:

$$\det(\lambda I - A) = \det \begin{bmatrix} \lambda & -1 \\ 2 & \lambda + 3 \end{bmatrix} = \lambda^2 + 3\lambda + 2 = 0$$

Resolviendo tal ecuación cuadrática se obtienen las raíces y por ende autovectores  $\lambda_1 = -1$  y  $\lambda_2 = -2$ . Se procede entonces a encontrar los auto vectores  $\vec{x}_1$  y  $\vec{x}_2$ . Para el auto-vector  $\vec{x}_1$ :

$$(\lambda_1 I - A)\vec{x}_1 = 0 \Rightarrow \begin{bmatrix} \lambda_1 & -1 \\ 2 & \lambda_1 + 3 \end{bmatrix} \vec{x}_1 = 0 \Rightarrow \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

Resolviendo tal sistema de ecuaciones, se obtiene que el auto-vector  $\vec{x}_1$  viene dado por:

$$\vec{x}_1 = k_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

y de manera similar, se obtiene el auto-vector  $\vec{x}_2$ :

$$\vec{x}_2 = k_2 \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Sin embargo, la notación anterior de los auto-vectores no está normalizada, por lo que usualmente se expresan de forma normalizada:

$$\hat{x}_1 = k_1 \frac{\vec{x}_1}{\|\vec{x}_1\|_2} = k_1 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

$$\hat{x}_2 = k_2 \frac{\vec{x}_2}{\|\vec{x}_2\|_2} = k_2 \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

En MATLAB los auto-vectores y auto-valores pueden calcularse como sigue:

```
>> A=[0 1;-2 -3]
>> [v,d] = eig(A)
```

Dada la complejidad de resolver el determinante para matrices grandes, se implementan otros métodos numéricos para calcular los auto-valores y auto-vectores.

Detallando más la interpretación geométrica de los auto-vectores, recordamos la igualdad en términos de la matriz cuadrada  $A \in \mathbb{R}^{n \times n}$ , los auto-valores y auto-vectores  $A\vec{v} = \lambda\vec{v}$ , la matriz  $A$  actúa como una transformación del auto-vector  $\vec{v}$ , la cual “envía” el vector a un nuevo punto del espacio, como se observa en la Figura 5. Recuerde que la multiplicación  $A\vec{x}$  realiza una combinación lineal de los componentes de  $\vec{v}$ :

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} a_{1,1}v_1 + a_{1,2}v_2 \\ a_{2,1}v_1 + a_{2,2}v_2 \end{bmatrix} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \end{bmatrix} v_1 + \begin{bmatrix} a_{1,2} \\ a_{2,2} \end{bmatrix} v_2$$

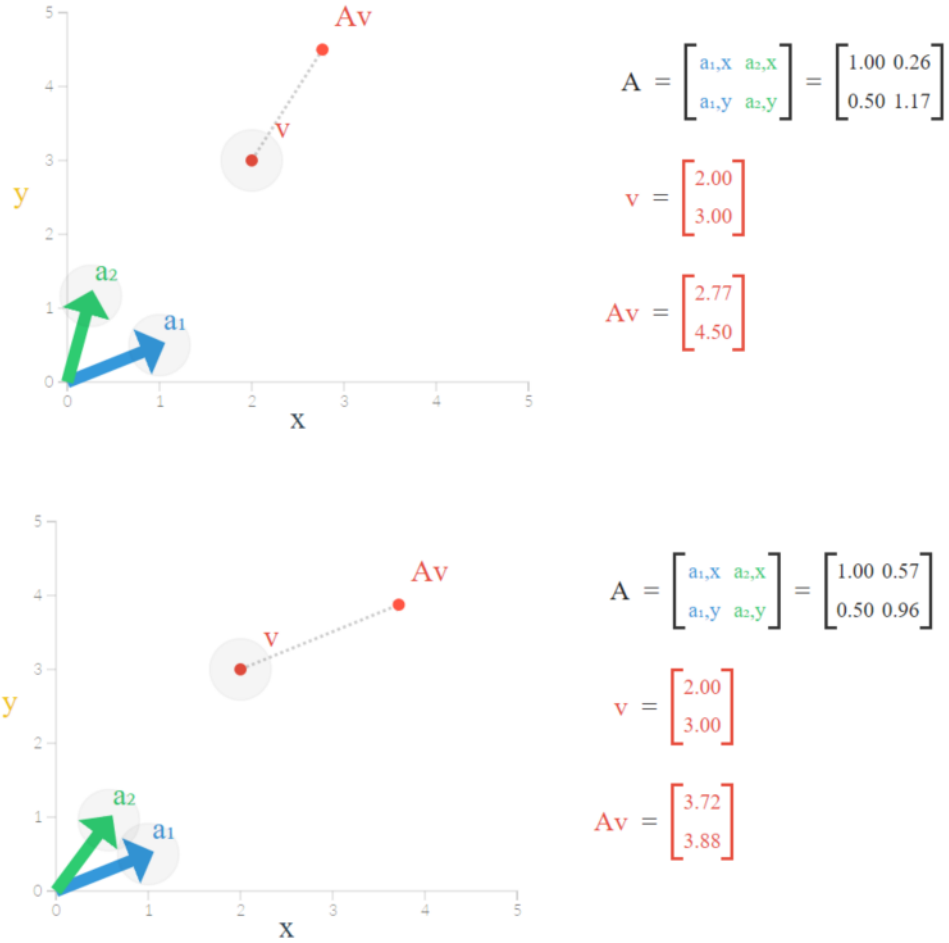


Figura 5: Transformación  $A\vec{x}$ , tomado de <http://setosa.io/ev/eigenvectors-and-eigenvalues/>.

Observe entonces que la ecuación de los auto-vectores  $A\vec{v} = \lambda\vec{v}$  corresponde a los vectores que transformados por la matriz  $A$ , son escalados por su auto-valor correspondiente  $\lambda$ , es decir, son los vectores que **conservan su dirección al ser transformados por la matriz  $A$** . Esto se puede verificar fácilmente, si es posible dibujar una línea recta entre los puntos  $(0,0)$ , y el final de los vectores  $\vec{v}$  y  $A\vec{v}$ . En la Figura 5, el primer caso puede corresponder a un auto-vector, mientras que el segundo, posiblemente no, pues la dirección del mismo es modificada al transformarse por la matriz  $A$ . Además, es fácil ver que todos los vectores con la misma dirección de un auto-vector  $\vec{v}$ , son también auto-vectores de tal matriz  $A$ , como se observa en la Figura 6, donde el auto-vector  $\vec{s}_1$  es colinear con el vector  $\vec{v}$ . Finalmente, es importante notar que

un auto-vector  $\vec{s}_1$  con su auto-valor  $\lambda_1 < 1$  denota una transformación  $A$  que “encoge” al auto-vector  $\vec{s}_1$ , y si  $\lambda_1 > 1$  más bien lo “alarga”, como se ve en la Figura 6. Un conjunto o sub-conjunto de los auto-vectores, conforman lo que se llama un auto-espacio.

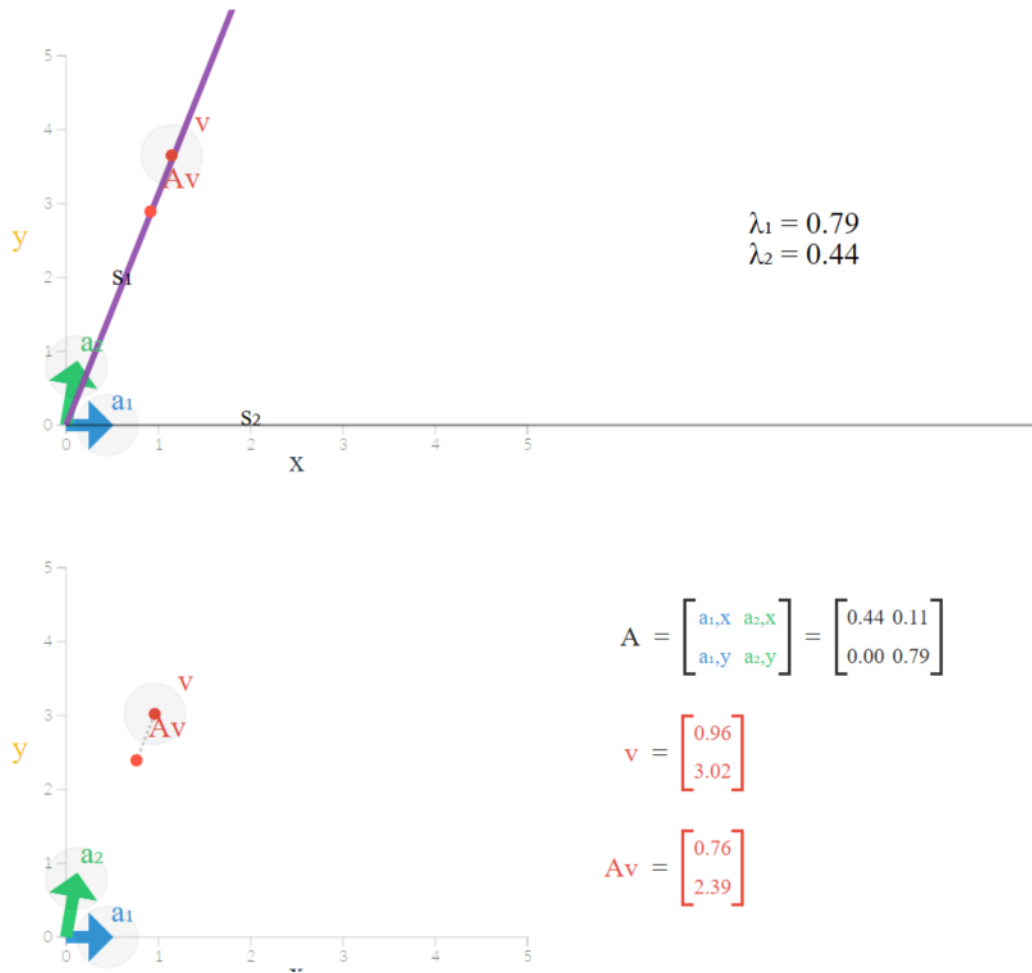


Figura 6: Auto-vectores  $\vec{s}_1$  y  $\vec{s}_2$ , tomado de <http://setosa.io/ev/eigenvectors-and-eigenvalues/>

Las siguientes son propiedades de los auto-valores y los auto-vectores, donde  $A \in \mathbb{R}^{n \times n}$ ,  $\vec{x} \in \mathbb{R}^n$  y  $\lambda \in \mathbb{R}$ :

- La traza de la matriz  $A$  es igual a la suma de sus auto-valores:

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i.$$



- El determinante de la matriz  $A$  es igual al producto de sus auto-valores:

$$\det(A) = \prod_{i=1}^n \lambda_i$$

- **El rango de la matriz  $A$  es igual al número de auto-valores no nulos.**  
Lo anterior tiene mucho sentido, pues si por ejemplo, con una matriz  $A \in \mathbb{R}^{2 \times 2}$  con una columna combinación lineal de la otra, la contribución independiente de una de ellas en  $A$  como transformación lineal es nula, como se ilustra en la Figura 7.

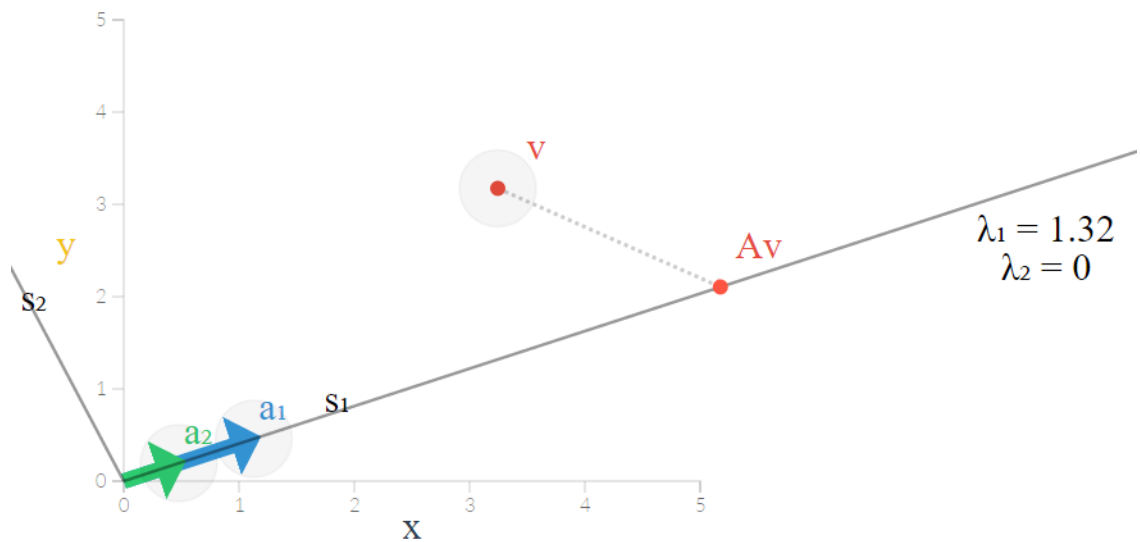


Figura 7: Auto-vectores de una matriz con columnas que son combinación lineal de la otra.

- Si  $A$  es no-singular (invertible), entonces  $1/\lambda_i$  es un auto-valor de  $A^{-1}$  con su auto-vector asociado  $\vec{x}_i$ , por lo que entonces  $A^{-1}\vec{x}_i = (1/\lambda_i)\vec{x}_i$ .
- Los auto-valores de una matriz diagonal  $D = \text{diag}(d_1, \dots, d_n)$  corresponden a tales entradas diagonales  $d_1, \dots, d_n$ .
- **Para una matriz simétrica  $A \in \mathbb{R}^{n \times n}$  existen  $\{v_1, v_2, \dots, v_n\}$  auto-vectores mutuamente ortogonales**, por ejemplo:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

Es usual que para expresar más fácilmente los auto-vectores y auto-valores en una sola ecuación:

$$AX = X\Lambda$$

con la matriz  $X \in \mathbb{R}^{n \times n}$  la cual agrupa los auto-vectores por columnas como:

$$X = \begin{bmatrix} | & | & \cdots & | \\ \vec{x}_{1,:} & \vec{x}_{2,:} & \cdots & \vec{x}_{n,:} \\ | & | & \cdots & | \end{bmatrix}$$

y la matriz  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  con los auto-valores de la transformación  $A$  en su diagonal. Note que los auto-vectores de  $A$  pueden ser linealmente dependientes, por lo que solo si los vectores columna son linealmente independientes se puede escribir:

$$A = X \Lambda X^{-1}$$

si lo anterior es posible de escribir, se dice que la matriz  $A$  es **diagonalizable**.

### 1.1.8. Auto-valores y auto-vectores de matrices simétricas

Las siguientes son propiedades de los auto-valores y auto-vectores para cualquier matriz simétrica  $A \in \mathbb{S}^n$ ,  $A = A^T$ :

- Los auto-valores de la matriz son siempre reales.
- Los auto-vectores son ortonormales, es decir, la matriz  $A X = X \Lambda$  es una matriz ortogonal, por lo que la matriz con los auto-vectores  $X$  se denota como  $U$ , por ello:  $A = U \Lambda U^{-1}$ , y recordando que la transpuesta de una matriz inversa equivale a la transpuesta, se tiene que:

$$A = U \Lambda U^T$$

## 2. Probabilidades

### 2.1. Funciones de densidad de probabilidad y de distribución

Una función de densidad de probabilidad con variable continua  $x$ ,  $p_X(x) \equiv p(X=x)$  o más resumido como  $p(x)$  se define de tal manera si para un intervalo muy pequeño  $\delta x \rightarrow 0$ , la probabilidad de que la variable aleatoria  $x$  esté en un intervalo  $(x, x + \delta x)$  es también infinitamente pequeña:  $p(x) \delta x \rightarrow 0$ .

Toda función de densidad de probabilidad debe cumplir las siguientes condiciones básicas:

$$0 \leq p(x) \leq 1$$

$$\int_{-\infty}^{\infty} p(x) \mathbf{d}x = 1.$$

### Función de distribución

La probabilidad de que  $x$  se encuentre en el intervalo  $(-\infty, z)$  está dada por la función de distribución acumulativa, o función de distribución  $P(z)$  definida como:

$$P_X(z) = P(z) = \int_{-\infty}^z p(x) dx$$

lo que implica que

$$P'(x) = p(x)$$

, como se muestra en la Figura ??.

Si la función de densidad  $p(x)$  se define como discreta también referida como función de , de manera similar se tienen las siguientes propiedades:

$$0 \leq p[x] \leq 1$$

$$\sum_{x=0}^{\infty} p[x] = 1.$$

$$P[z] = \sum_{x=0}^z p[x]$$

Las siguientes son propiedades de la función de distribución  $P(z)$ :

- $0 \leq P(z) \leq 1.$
- $\lim_{z \rightarrow \infty} P(z) = 1$
- $\lim_{z \rightarrow -\infty} P(z) = 0$

La función de densidad de probabilidad con múltiples variables  $x_1, \dots, x_D$

se denota de forma compacta con el vector  $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix} = (x_1, \dots, x_D)$  como la

función de densidad de probabilidad conjunta  $p(\vec{x}) = p(x_1, \dots, x_D)$ , por lo que la probabilidad de que  $\vec{x}$  se encuentre en un volumen infinitesimal  $\delta\vec{x}$  está dado por  $p(\vec{x}) \delta\vec{x}$  y cumple también las dos propiedades básicas:

$$p(\vec{x}) \geq 0$$

$$\int_{-\infty}^{\infty} p(\vec{x}) d\vec{x} = 1.$$

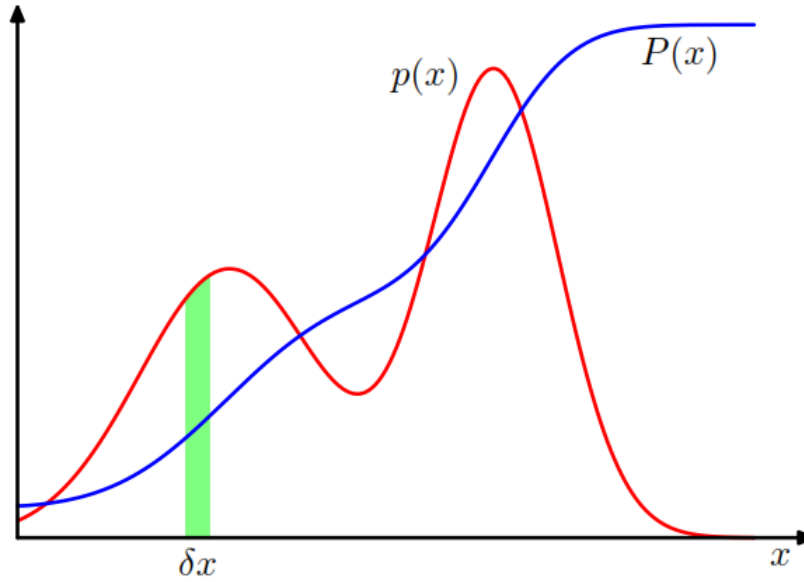


Figura 8: Gráfica de las funciones de densidad y de distribución[?].

### 2.1.1. Esperanza, varianza y covarianza de una función de densidad de probabilidad

El cálculo de momentos estadísticos es una operación muy utilizada en el reconocimiento de patrones, pues permite reducir la dimensionalidad de datos, recabando características importantes como el valor esperado o la varianza.

Se define entonces la **esperanza o el valor esperado** de una variable aleatoria  $X$  como la sumatoria pesada por la probabilidad de cada valor  $x$  que puede tomar tal variable aleatoria (a la izquierda en el caso de ser una variable aleatoria continua, a la derecha discreta):

$$\mu_X = \mathbb{E}[X] = \int x p(x) dx \quad \mu_X = \mathbb{E}[X] = \sum_x x p[x]$$

En el caso de una función discreta  $h[u]$  la cual acumula el valor  $x$  generado registrado para el experimento  $u$  para un total de  $N$  experimentos, generados a partir de la variable aleatoria  $X$ , la esperanza está dada por:

$$\mathbb{E}[X] \cong \frac{1}{N} \sum_{u=1}^N h[u]. \quad (3)$$

Desde un punto de vista frecuentista, la aproximación de  $\mathbb{E}[X]$  mejora, a medida que  $N \rightarrow \infty$ .

Las siguientes son propiedades de la esperanza:

- Si  $a$  es un escalar, tal que  $a \in \mathbb{R}$ , se tiene que:  $\mathbb{E}[a] = a$ .
- Homogeneidad:  $\mathbb{E}[aX] = a \mathbb{E}[X]$ .
- Superposición:  $\mathbb{E}[g(X) + f(X)] = \mathbb{E}[g(X)] + \mathbb{E}[f(X)]$ .

### Varianza

La varianza de una variable aleatoria  $X$  es una medida de la “concentración” o dispersión de los datos alrededor de la media. Formalmente la varianza de una variable aleatoria  $X$  se define como:

$$\sigma_X^2 = \text{var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2],$$

lo cual equivale a (por la linealidad de la esperanza y su propiedad de equivalencia en constantes):

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ \Rightarrow \text{var}[X] &= (\mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2]) \\ \Rightarrow \text{var}[X] &= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\ \Rightarrow \text{var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

Propiedades de la varianza:

- Si  $a$  es un escalar, tal que  $a \in \mathbb{R}$ , se tiene que:  $\text{Var}[a] = 0$ .
- Multiplicación por escalar de la entrada:  $\text{Var}[aX] = a^2 \text{Var}[X], \forall a \in \mathbb{R}$ .

En el caso de una función discreta  $h[u]$  la cual acumula el valor  $x$  generado registrado para el experimento  $u$  para un total de  $N$  experimentos, generados a partir de la variable aleatoria  $X$ , la varianza está dada por:

$$\sigma_X^2 \cong \frac{1}{N-1} \sum_{u=1}^N (h[u] - \mu_X)^2. \quad (4)$$

El estimador anterior se dice que no es “sesgado” si la media real no es conocida. Si la media real de la población es conocida, se normaliza por  $N$  (puede ver la demostración de lo anterior en <http://www.visiondummy.com/2014/03/divide-variance-n-1/>).

### Ejemplo

Calcule la media y la varianza de una variable aleatoria uniforme  $X$  con una función de densidad  $p(x) = 1, \forall x \in [0, 1]$ , 0 de otro modo:

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x p(x) dx = \int_0^1 x dx = \frac{1}{2} \\ \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 p(x) dx = \int_0^1 x^2 dx = \frac{1}{3} \\ \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\end{aligned}$$

## Covarianza

Para dos variables aleatorias  $X$  e  $Y$  se la covarianza como:

$$\Sigma_{X,Y} = \text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

y mide la variación conjunta de tales variables aleatorias. Para el caso de contar con arreglos de muestras  $h[u]$  y  $g[u]$  para las variables aleatorias  $X$  e  $Y$  respectivamente, se tiene que la covarianza de tales variables aleatorias está dada por:

$$\Sigma_{X,Y} \cong \frac{1}{N-1} \sum_{u=1}^N (h[u] - \mu_X)(g[u] - \mu_Y).$$

La matriz de covarianza para  $n$  variables aleatorias  $X_1, X_2, \dots, X_n$  se define como:

$$\Sigma = \begin{bmatrix} \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_1 - \mathbb{E}[X_1])] & \dots & \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_n - \mathbb{E}[X_n])] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[(X_n - \mathbb{E}[X_n])(X_1 - \mathbb{E}[X_1])] & \dots & \mathbb{E}[(X_n - \mathbb{E}[X_n])(X_n - \mathbb{E}[X_n])] \end{bmatrix},$$

observe que en la diagonal de la matriz  $\Sigma$  (entrada  $\Sigma_{i,i}$ ) se tiene que

$$\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_i - \mathbb{E}[X_i])] = \sigma_{X_i}^2,$$

por lo que entonces la matriz de covarianza se puede reescribir como:

$$\Sigma = \begin{bmatrix} \sigma_{X_1}^2 & \dots & \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_n - \mathbb{E}[X_n])] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[(X_n - \mathbb{E}[X_n])(X_1 - \mathbb{E}[X_1])] & \dots & \sigma_{X_n}^2 \end{bmatrix}.$$

Además, la matriz de covarianza  $\Sigma$  presenta la propiedad de ser simétrica, puesto que  $\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \mathbb{E}[(X_j - \mathbb{E}[X_j])(X_i - \mathbb{E}[X_i])] \Rightarrow \Sigma_{X_i, X_j} = \Sigma_{X_j, X_i}$ .

### Ejemplo

Suponga que se desea encontrar la matriz de covarianza para tres variables aleatorias  $X_1, X_2$  y  $X_3$ , para las cuales se han recabado los siguientes arreglos de muestras para  $N = 4$  experimentos, respectivamente:

$$\begin{aligned} h_1 &= [2 \quad 4 \quad 6 \quad 8] \\ h_2 &= [4 \quad 8 \quad 12 \quad 16] \\ h_3 &= [12 \quad 10 \quad 5 \quad 9] \end{aligned}$$

En términos de muestras se tienen 4 muestras

$$U = \{\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4\} = \begin{bmatrix} | & | & | & | \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 8 & 12 & 16 \\ 12 & 10 & 5 & 9 \end{bmatrix}$$

con  $u_i \in \mathbb{R}^3$ , donde cada dimensión es una variable aleatoria, y  $U \in \mathbb{R}^{3 \times 4}$ .

Observe en estos datos, que la dimensión 1 y 2 son combinación lineal para todas las muestras, por lo que la covarianza de ambas dimensiones debe ser alta, no así la dimensión 1 con la 3 o la 2 con la 3. Además

Se procede entonces a calcular las entradas  $\Sigma_{X_1, X_2}$ ,  $\Sigma_{X_1, X_3}$  y  $\Sigma_{X_2, X_3}$ , además de los valores de la diagonal  $\sigma_{X_1}^2$ ,  $\sigma_{X_2}^2$  y  $\sigma_{X_3}^2$ , teniendo en cuenta que  $\mu_{X_1} = 5$ ,  $\mu_{X_2} = 10$  y  $\mu_{X_3} = 9$ :

$$\begin{aligned}\Sigma_{X_1, X_2} &= \frac{1}{4-1} ((5-2)(10-4) + (5-4)(10-8) + (5-6)(10-12) + (5-8)(10-16)) \\ \Sigma_{X_1, X_3} &= \frac{1}{4-1} ((5-2)(9-12) + (5-4)(9-10) + (5-6)(9-5) + (5-8)(9-9)) \\ \Sigma_{X_2, X_3} &= \frac{1}{4-1} ((10-4)(9-12) + (10-8)(9-10) + (10-12)(9-5) + (10-16)(9-9)) \\ \sigma_{X_1}^2 &= \frac{1}{4-1} ((5-2)^2 + (5-4)^2 + (5-6)^2 + (5-8)^2) \\ \sigma_{X_2}^2 &= \frac{1}{4-1} ((10-4)^2 + (10-8)^2 + (10-12)^2 + (10-16)^2) \\ \sigma_{X_3}^2 &= \frac{1}{4-1} ((9-12)^2 + (9-10)^2 + (9-5)^2 + (9-9)^2)\end{aligned}$$

lo cual desarrollado corresponde a:

$$\begin{aligned}\Sigma_{X_1, X_2} &= \frac{1}{3} (3 \cdot 6 + 1 \cdot 2 + -1 \cdot -2 + -3 \cdot -6) = \frac{40}{3} = 13,333 \\ \Sigma_{X_1, X_3} &= \frac{1}{3} (3 \cdot -3 + 1 \cdot -1 + -1 \cdot 4 + -3 \cdot 0) = -\frac{14}{3} = -4,667 \\ \Sigma_{X_2, X_3} &= \frac{1}{3} (6 \cdot -3 + 2 \cdot -1 + -2 \cdot 4 + -6 \cdot 0) = -\frac{28}{3} = -9,333 \\ \sigma_{X_1}^2 &= \frac{1}{3} (9 + 1 + 1 + 9) = \frac{20}{3} = 6,667 \\ \sigma_{X_2}^2 &= \frac{1}{3} (36 + 4 + 4 + 36) = \frac{80}{3} = 26,667 \\ \sigma_{X_3}^2 &= \frac{1}{3} (9 + 1 + 16 + 0) = \frac{14}{3} = 4,667.\end{aligned}$$

Por lo que se obtiene la matriz de covarianza:

$$\Sigma = \begin{bmatrix} \frac{20}{3} & \frac{40}{3} & -\frac{14}{3} \\ \frac{40}{3} & \frac{80}{3} & -\frac{28}{3} \\ -\frac{14}{3} & -\frac{28}{3} & \frac{14}{3} \end{bmatrix} = \begin{bmatrix} 6,667 & 13,333 & -4,667 \\ 13,333 & 26,666 & -9,333 \\ -4,667 & -9,333 & 4,667 \end{bmatrix}.$$

Observe que la covarianza más alta es  $\Sigma_{X_1, X_2}$  pues las dimensiones 1 y 2, cuyos datos son generados por las variables aleatorias  $X_1$  y  $X_2$ , pues los datos siempre covarían positivamente para todas las muestras, en ambas dimensiones. En los otros dos casos, donde la covarianza es negativa, ello denota que cuando en una dimensión los datos varían en una dirección, en la otra los datos varían en dirección contraria.

En MATLAB lo anterior se puede implementar como sigue:

```
U = [2 4 6 8; 4 8 12 16; 12 10 5 9];
%Cada columna son los datos de una dimension
matCov = cov(U');
```

También la matriz de covarianza se puede escribir, para un conjunto de muestras  $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m\}$ , con  $\vec{x}_i \in \mathbb{R}^{n \times 1}$ , como:

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

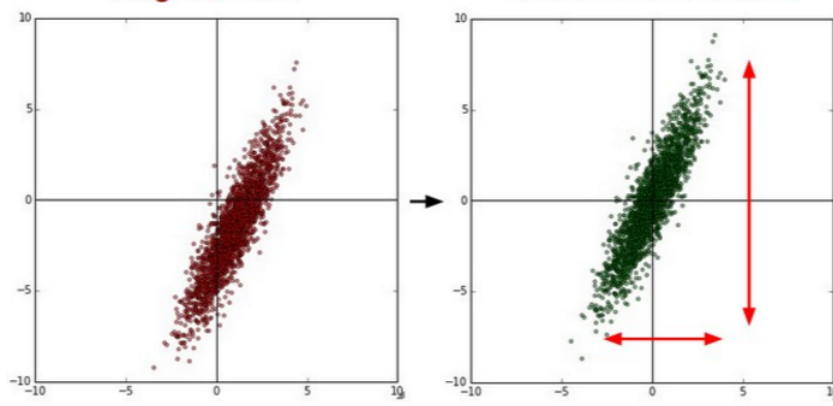


Figura 9: Modificación del origen de los datos, sustrayendo la muestra media  $\vec{\mu} \in \mathbb{R}^{n \times 1}$ .

donde  $\vec{\mu} \in \mathbb{R}^{n \times 1}$  es la muestra promedio del conjunto de datos  $X$  (donde cada componente es el valor medio de cada dimensión).

Definiendo a  $\vec{\delta}_i = \vec{x}_i - \vec{\mu}$ , se reescribe la ecuación anterior como:

$$\Sigma = \frac{1}{m-1} \sum_{i=1}^m \vec{\delta}_i \vec{\delta}_i^T$$

Observe que lo anterior se refiere a crear *muestras centradas en la media de los datos*. Esto tiene un efecto en los signos de los valores, por ejemplo todos los valores originalmente pudieron ser positivos, pero al realizar la sustracción de la muestra media, muchos de las muestras en distintas dimensiones se harán negativos. Lo anterior se ilustra en la Figura 9.

Para el caso en que  $\vec{x}_i \in \mathbb{R}^2$  se tiene que la matriz de covarianza está dada por:

$$\Sigma = \begin{bmatrix} \frac{1}{m} \sum_{i=1}^m (x_1 - \mu_1)^2 & \frac{1}{m} \sum_{i=1}^m (x_1 - \mu_1)(x_2 - \mu_2) \\ \frac{1}{m} \sum_{i=1}^m (x_2 - \mu_2)(x_1 - \mu_1) & \frac{1}{m} \sum_{i=1}^m (x_2 - \mu_2)^2 \end{bmatrix}$$

$$\Rightarrow \Sigma = \begin{bmatrix} \sigma_1^2 & \frac{1}{m} \sum_{i=1}^m (x_1 - \mu_1)(x_2 - \mu_2) \\ \frac{1}{m} \sum_{i=1}^m (x_2 - \mu_2)(x_1 - \mu_1) & \sigma_2^2 \end{bmatrix}.$$

Recordemos que  $\sigma_1^2$  define la dispersión de los datos en la primer dimensión, y  $\sigma_2^2$  en la segunda dimensión. Para la matriz de covarianza de ejemplo, las entradas  $\Sigma_{2,1}$  y  $\Sigma_{1,2}$  equivalentes denotan la covarianza de las variables aleatorias  $X_1$  y  $X_2$ . La Figura 10 en la parte izquierda, grafica un caso en el que la covarianza de dos variables aleatorias  $X_1$  y  $X_2$  es positiva, lo que denota una pendiente creciente en la dispersión de los datos, y a la derecha, el caso en el que la covarianza es negativa, haciendo una pendiente negativa en la dispersión de los datos. En caso de no existir ninguna pendiente u orientación en la



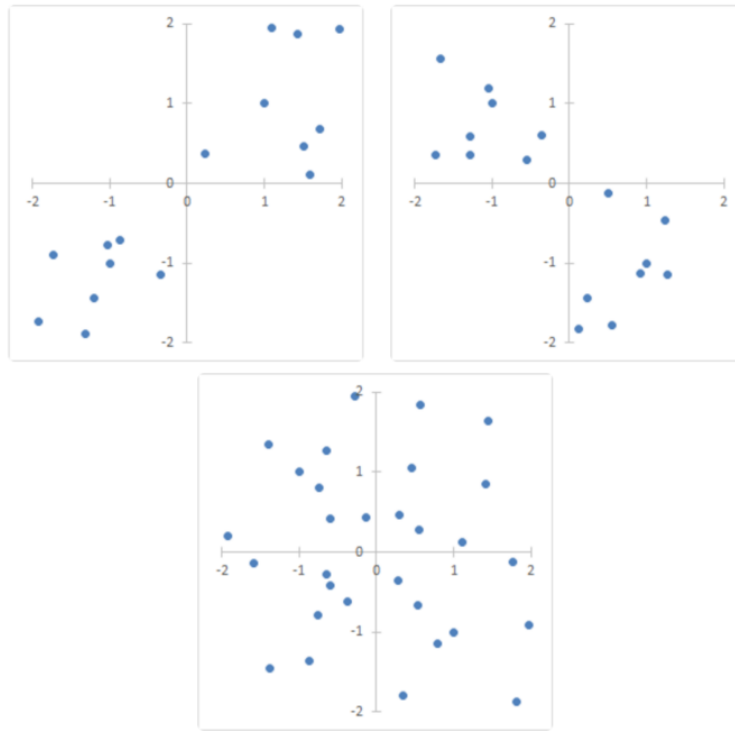


Figura 10: Esquina superior izquierda; covarianza positiva, esquina superior derecha; covarianza negativa, gráfica inferior; covarianza tendiendo a cero.

dispersión de los datos, la covarianza entre ambas variables aleatorias tiende a cero, como muestra también la gráfica inferior en la Figura 10.

Un ejemplo sencillo para entender mejor la covarianza entre dos variables aleatorias  $X_1$  y  $X_2$ , es darse el caso de una imagen compuesta únicamente por dos píxeles, de forma que  $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ , con dos funciones de densidad  $p(X_1 = x_1)$  y  $p(X_2 = x_2)$ . Si tras varias muestras los píxeles en ambas posiciones tienden a ser similares (por ejemplo negro y negro), la covarianza será alta. Si en cambio, los píxeles tienden a tener valores opuestos (por ejemplo blanco y negro), su covarianza es negativa. Finalmente, si los píxeles tienen algunas veces valores opuestos y en otros similares, se eliminarán las contribuciones en el cálculo de la covarianza, por lo que su valor tenderá a ser nulo.

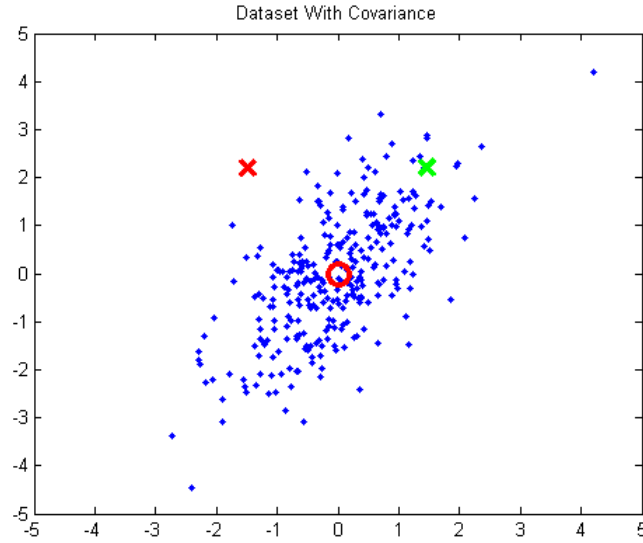


Figura 11: Datos con covarianza positiva.

## 2.2. El efecto de la correlación en la distancia, auto-valores y auto-vectores de la matriz de covarianza

Para entender mejor el efecto de la correlación o covarianza en el cálculo de la distancia, tómese los datos con covarianza positiva de la Figura 11. El dato con la «O» roja corresponde al centroide o valor medio  $\vec{\mu} \in \mathbb{R}^2$ , y las «X» roja y verde corresponde a dos puntos  $\vec{x}_1 \in \mathbb{R}^2$  y  $\vec{x}_2 \in \mathbb{R}^2$  respectivamente, a igual distancia Euclidiana de tal centroide.

Observando tal gráfica, se observa que  $\vec{x}_1$  (X roja) es menos probable que pertenezca al clúster que el vector  $\vec{x}_2$  («X» verde), inclusive tomando en cuenta la varianza en ambas dimensiones, pues es la covarianza la que altera la verosimilitud del punto  $\vec{x}_1$  («X» roja).

Queda claro entonces que es necesario tomar en cuenta la covarianza de los datos. Tal información se encuentra en la matriz de covarianza. Para entender mejor como es que la distancia de Mahalanobis toma en cuenta tal distancia, se realizará una transformación para remover la correlación y la varianza.

La remoción de la correlación y la covarianza se refiere a una técnica llamada Blanqueamiento de Componentes Principales (*PCA whitening*). Para realizarlo, se hacen los pasos básicos del ACP:

1. Calcular la matriz de covarianza  $\Sigma$ .
2. Calcular los auto-vectores  $\vec{v}_1$  y  $\vec{v}_2$  de la covarianza  $\Sigma$ , los cuales constituyen los componentes principales de los datos. El primer componente principal  $\vec{v}_1$  (con mayor auto valor  $\lambda_1$ ) tiene como dirección la dirección

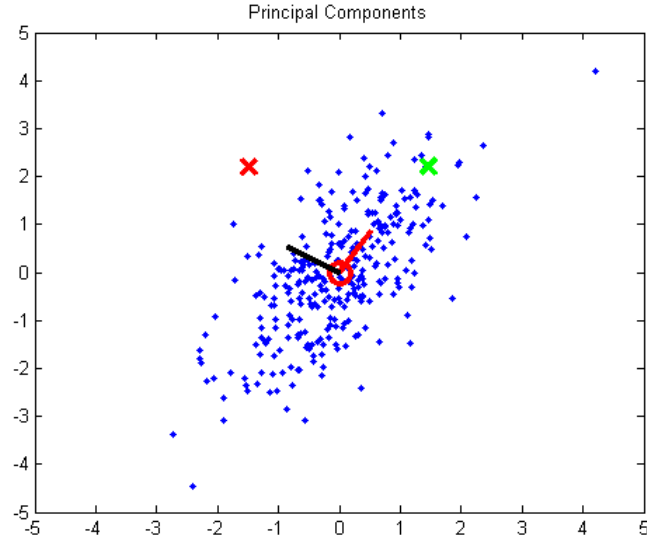


Figura 12: Componentes principales de los datos.

con mayor varianza de los datos, y  $\vec{v}_2$  la dirección con menor varianza, como muestra la Figura 12.

### 3. Proyectar los datos en el nuevo espacio.

Usando tales auto-vectores  $\vec{v}_1$  y  $\vec{v}_2$ , es posible rotar los datos para que la dirección de la máxima varianza esté alineada con el eje  $x$ , y la segunda dirección de mayor varianza esté alineada con el eje  $y$ . Para realizar tal rotación, se proyectan los datos en los dos componentes principales, de forma que para un punto con la media sustraída  $\vec{u}_i = \vec{x}_i - \vec{\mu}$ , se realice el cálculo del vector proyectado  $\vec{u}'_i$  usando como base los auto-vectores  $\vec{v}_1$  y  $\vec{v}_2$ . Ese cálculo está dado para cada componente de  $\vec{u}'_i$  como:

$$\vec{u}'_{i,1} = \vec{u}_i \cdot \vec{v}_1$$

$$\vec{u}'_{i,2} = \vec{u}_i \cdot \vec{v}_2.$$

La proyección en el espacio de los auto-vectores, resulta en la graficación de los puntos de la Figura 13.

Calculando la matriz de covarianza  $\Sigma$  de los datos rotados, se obtiene algo con la forma:

$$\Sigma = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

con  $a \neq 0$  y  $b \neq 0$ .

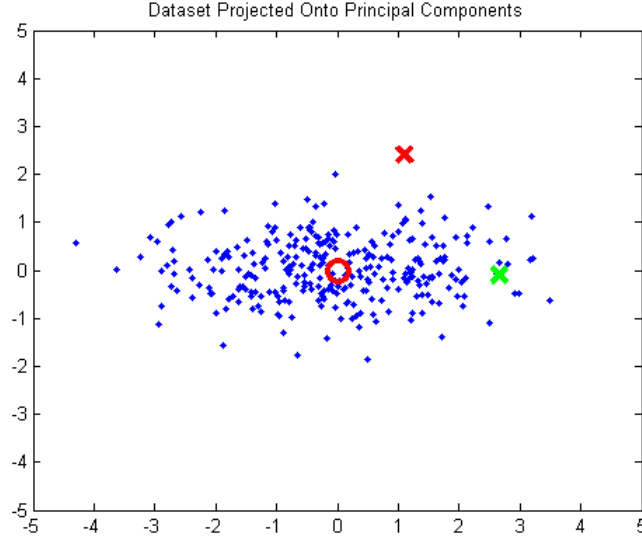


Figura 13: Proyección de los datos en los dos componentes principales.

El hecho de que no exista covarianza o correlación se refiere gráficamente a que los datos están igualmente distribuidos en los cuatro cuadrantes respecto al origen  $(0, 0)$ . Observe que aún los dos datos  $\vec{x}_1$  y  $\vec{x}_2$  presentan aún la misma distancia desde el origen. Sin embargo, las principales direcciones de la variación están ahora alineadas a los ejes  $x$  e  $y$ , por lo que podemos ahora normalizar los datos para que presenten una varianza unitaria, dividiendo los componentes por la raíz cuadrada de su varianza. Ello transforma el cluster de datos en una esfera, haciendo que efectivamente el punto  $\vec{x}_2$  («X» verde) esté más cerca del centroide de los datos, como muestra la Figura 14.

### 3. Análisis de componentes principales para rostros

El reconocimiento de caras propuesto en [?] 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  $\lambda$  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  $\lambda$  muy altos se presenta un caso de escalamiento muy alto para el auto vector  $\vec{v}$ . Para encontrar

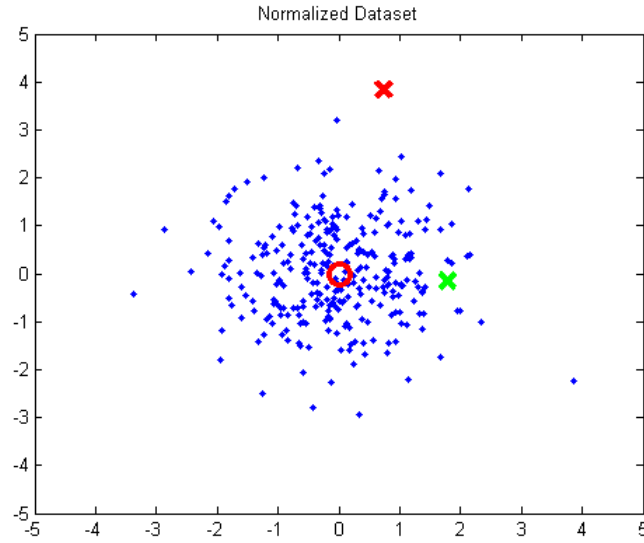


Figura 14: Datos normalizados.

los auto-valores  $\lambda$ , nos basamos en la suposición de que el sistema de ecuaciones anterior tiene infinitas soluciones, por lo que entonces el determinante de la ecuación anterior debe ser cero, es decir:

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

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  $\lambda_i$  es asociado a un auto-vector  $v_i$  que se obtiene al resolver la ecuación 3 con el valor encontrado de  $\lambda_i$ .

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 15 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 16.

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  $ev_3$  diagramado en la Figura 17.

Para encontrar tales «ejes», el ACP propone la descomposición en auto-valores y auto-vectores de la matriz de covarianza  $\Sigma$  asociada a la matriz de datos  $A$ . La matriz de covarianza  $\Sigma$  para datos de dimensionalidad  $n$  es de

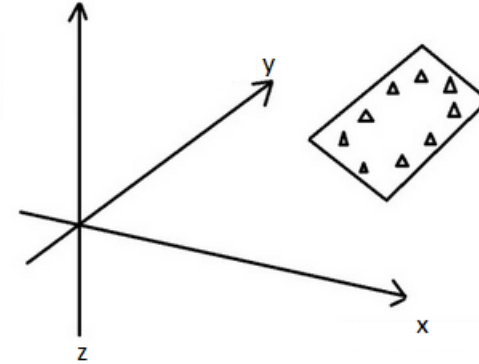


Figura 15: Datos de dimensionalidad 3.

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 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 18.

### 3.1. Auto-Caras para la clasificación

Lo propuesto en [?] 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 a la base de datos con imágenes 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:

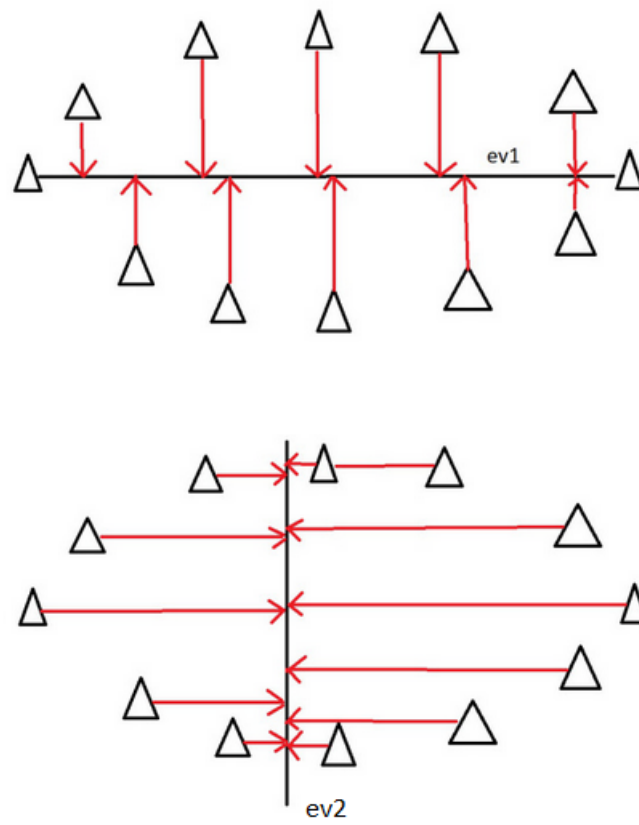


Figura 16: Ejes con mayor variación.

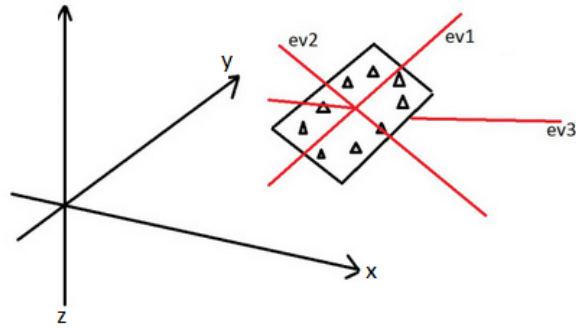


Figura 17: Autovectores de datos de dimensionalidad 3.

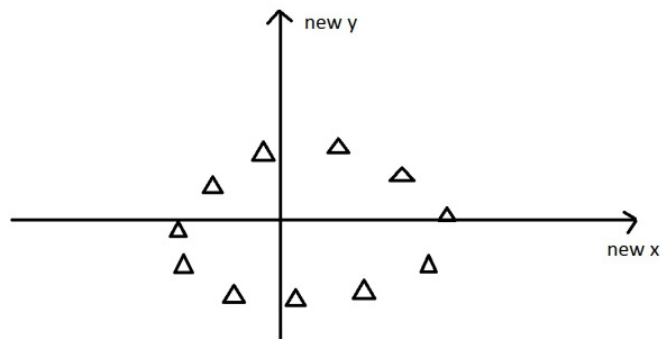


Figura 18: Nuevo espacio vectorial tomando los auto-vectores mas importantes.



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  $\Sigma$ .
  - 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  $\mu_i$ , con  $1 \leq i \leq s$  más cercana, usando una distancia euclidiana o  $L_2$ .
    - 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.

### 3.1.1. 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:

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

La matriz de diferencias respecto a la cara promedio, se define como  $D = [\vec{d}_1, \vec{d}_2, \dots, \vec{d}_m]$  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  $D \in \mathbb{R}^{n \times m}$ . 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  $640 \times 480$  pixeles,  $n = 307200$ ). Veremos posteriormente 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  $m \ll n$ .

### 3.1.2. Generación del nuevo espacio vectorial

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

$$\Sigma_v \vec{v}_i = \lambda_i \vec{v}_i \Rightarrow D^T D \vec{v}_i = \lambda_i \vec{v}_i, \quad (5)$$

sin embargo nos sigue interesando el cálculo de los auto-vectores para la matriz de covarianza  $\Sigma_w$ , pues **estos describen la covarianza en cada pixel de las imágenes**. Para hacerlo, se sigue la ecuación 5 la cual multiplicando a ambos lados por  $D$  se obtiene que:

$$D D^T D \vec{v}_i = \lambda_i D \vec{v}_i, \quad (6)$$

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

$$D D^T \vec{w}_i = \lambda_i \vec{w}_i, \quad (7)$$

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:

$$\Sigma_w \vec{w}_i = \lambda_i \vec{w}_i, \quad (8)$$

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  $\Sigma_w$ , 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  $E = \text{espacio Generado } \{\vec{w}_1, \dots, \vec{w}_{n'}\}$ . 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  $W$ .

### 3.2. 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  $t = i$ .

## Referencias