



Institución  
**Universitaria**  
Reacreditada en Alta Calidad

# VISIÓN ARTIFICIAL

CARLOS ANDRÉS MERA BANGUERO, PHD

[carlosmera@itm.edu.co](mailto:carlosmera@itm.edu.co)

PROGRAMA DE INGENIERÍA DE SISTEMAS

[www.itm.edu.co](http://www.itm.edu.co)

## SELECCIÓN DE CARACTERÍSTICAS

### Introducción

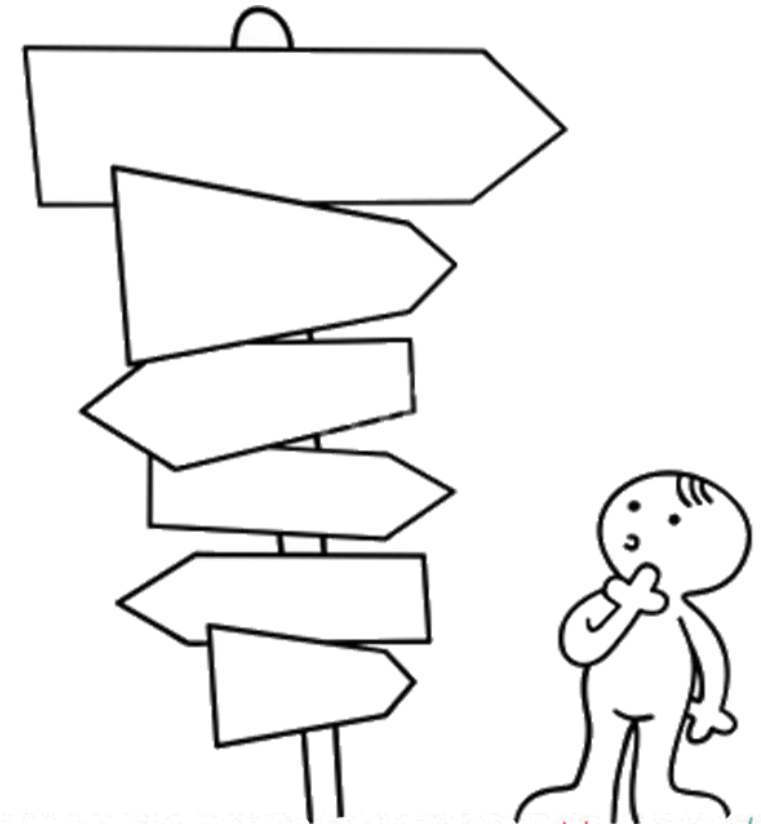
### El Problema de la Dimensionalidad

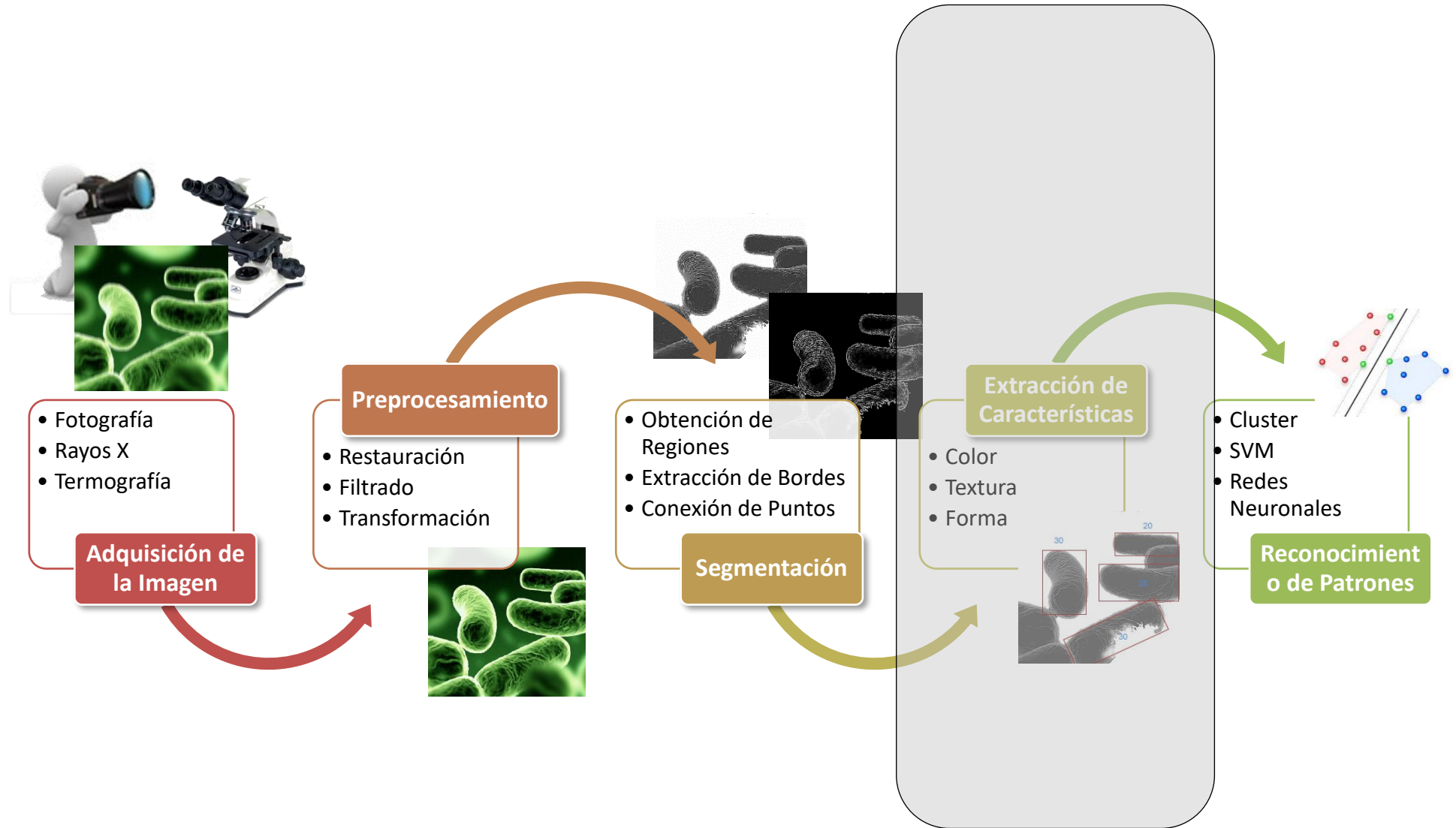
### Métodos de Selección de Características

- ✓ Selección Simple
- ✓ Búsqueda Exhaustiva
- ✓ Selección Hacia Adelante
- ✓ Selección Hacia Atrás
- ✓ Selección Branch and Bound

### Métodos de Extracción (o Transformación) de Características

- ✓ PCA





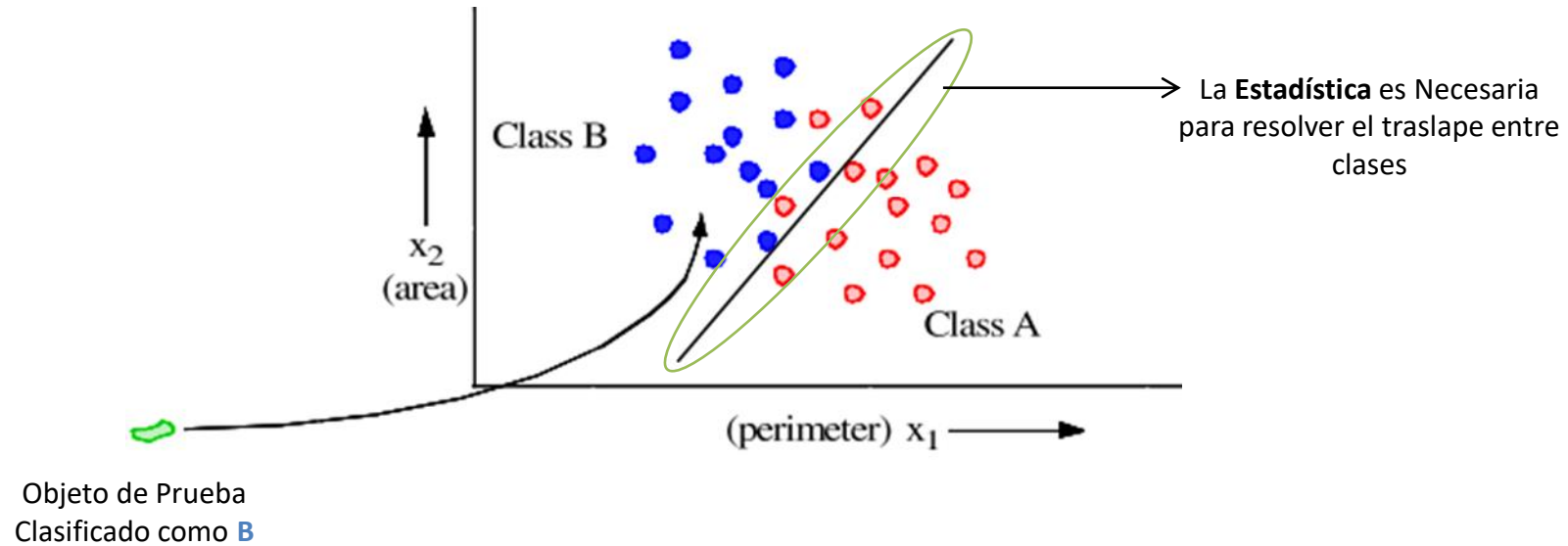
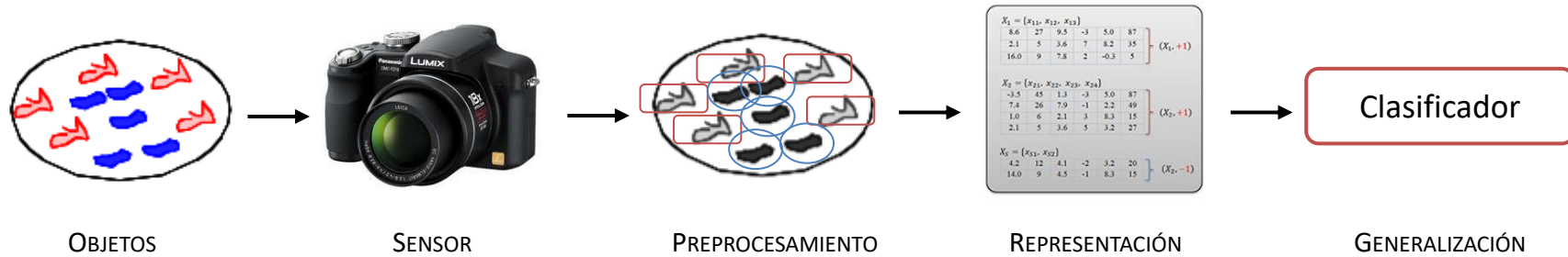


# INTRODUCCIÓN




# SELECCIÓN DE CARACTERÍSTICAS

## INTRODUCCIÓN





## INTRODUCCIÓN – REGLA DE BAYES

 El **Teorema de Bayes** expresa la probabilidad a posteriori de un evento aleatorio  $A$  dado  $x$  en términos de la distribución de probabilidad condicional y la probabilidad marginal.

 Para un problema de 2 clases:

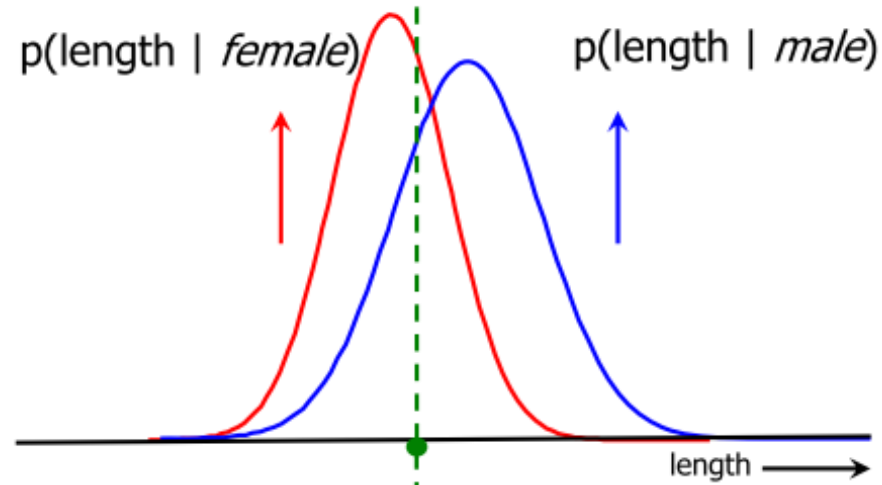
$$\begin{aligned} p(A|x) &> p(B|x) && \rightarrow A \text{ else } B \\ \frac{p(x|A) p(A)}{p(x)} &> \frac{p(x|B) p(B)}{p(x)} && \rightarrow A \text{ else } B \\ p(x|A) p(A) &> p(x|B) p(B) && \rightarrow A \text{ else } B \end{aligned}$$

  $S(x) = p(x|A)P(A) - p(x|B)P(B) > 0 \rightarrow A \text{ else } B$

 Para múltiples clases :  $Clase(x) = \arg \max_{\omega} (p(x|\omega)P(\omega))$

## INTRODUCCIÓN – REGLA DE BAYES

Cuál es el género de alguien con esta altura?



Bayes: { 
$$\begin{aligned} p(\text{female} | \text{length}) &= p(\text{length} | \text{female}) p(\text{female}) / p(\text{length}) \\ p(\text{male} | \text{length}) &= p(\text{length} | \text{male}) p(\text{male}) / p(\text{length}) \end{aligned}$$

## INTRODUCCIÓN – REGLA DE BAYES

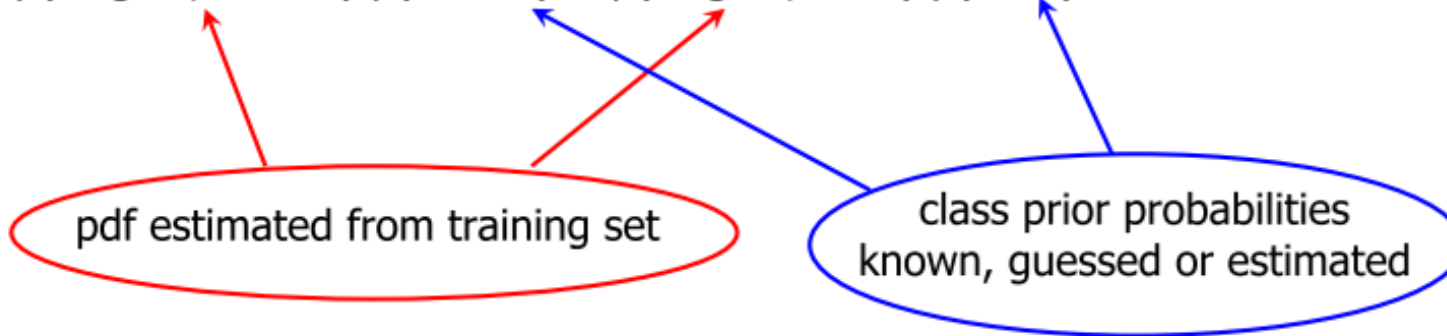
Regla de Clasificación de Bayes:

$$p(\textit{female} \mid \textit{length}) > p(\textit{male} \mid \textit{length}) \rightarrow \textit{female} \text{ else } \textit{male}$$

Bayes:

$$\frac{p(\textit{length} \mid \textit{female}) p(\textit{female})}{p(\textit{length})} > \frac{p(\textit{length} \mid \textit{male}) p(\textit{male})}{p(\textit{length})}$$

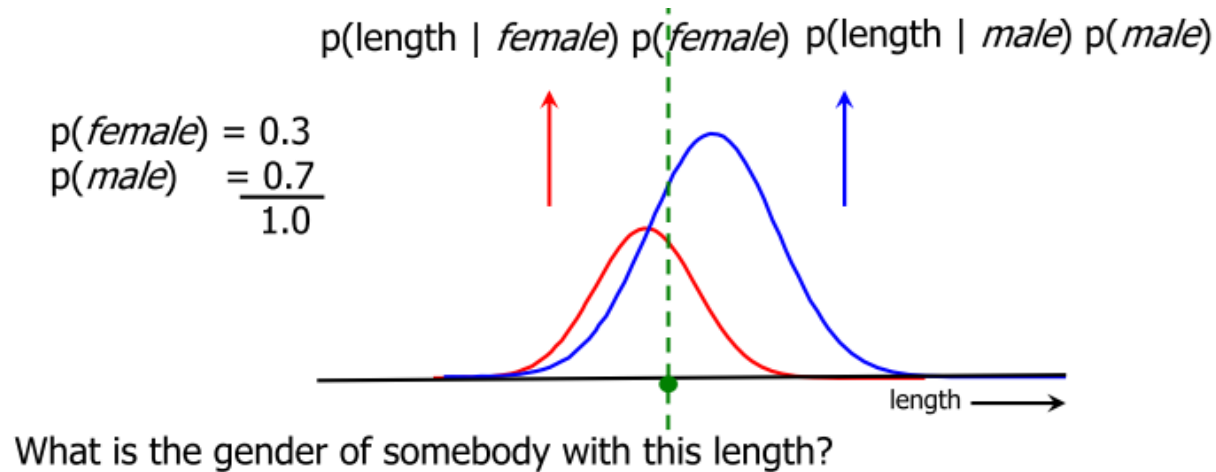
$$p(\textit{length} \mid \textit{female}) p(\textit{female}) > p(\textit{length} \mid \textit{male}) p(\textit{male}) \rightarrow \textit{female} \text{ else } \textit{male}$$





## INTRODUCCIÓN – REGLA DE BAYES

Cuál es el género de alguien con esta altura?



Bayes:  $\begin{cases} p(\text{female} | \text{length}) = p(\text{length} | \text{female}) p(\text{female}) / p(\text{length}) \\ p(\text{male} | \text{length}) = p(\text{length} | \text{male}) p(\text{male}) / p(\text{length}) \end{cases}$

¿Es la altura una buena característica para diferenciar hombres de mujeres?

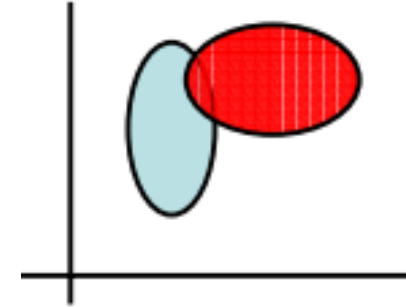


## BUENAS REPRESENTACIONES

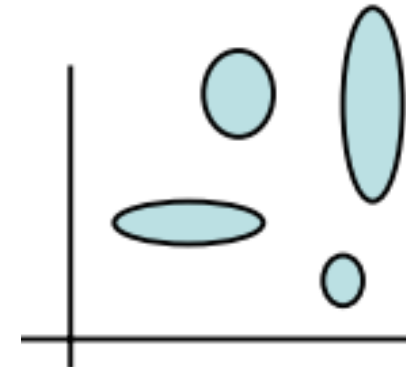
# SELECCIÓN DE CARACTERÍSTICAS

## 🏆 BUENAS REPRESENTACIONES

- 🎯 **Específicas a las Clases**  
Diferentes clases deberían estar representados en diferentes posiciones en el espacio de representación.

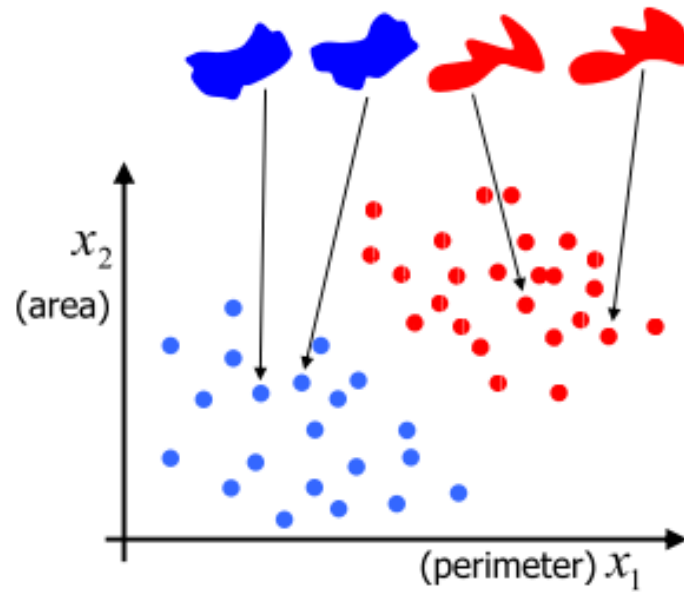


- 🎯 **Deben ser Compactas**  
Cada clase debe estar representada en un pequeño conjunto de dominios finitos.



# SELECCIÓN DE CARACTERÍSTICAS

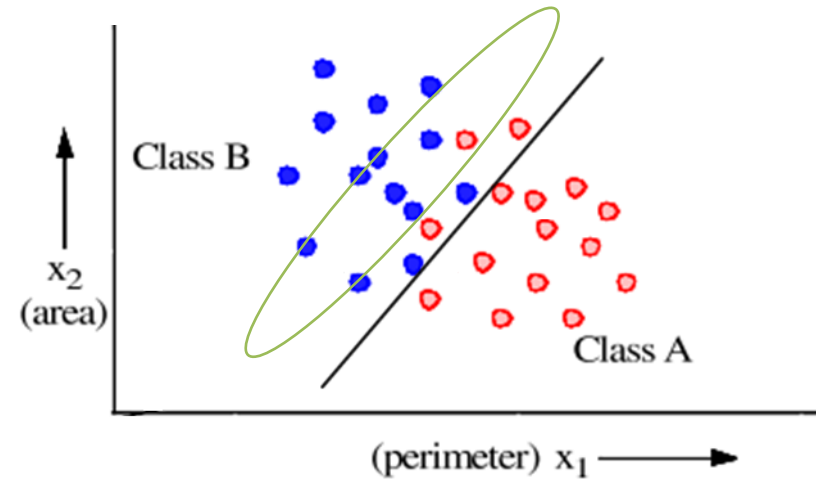
## BUENAS REPRESENTACIONES



Objetos **SIMILARES** están cerca  
y  
Objetos **DISÍMILES** son distantes

# SELECCIÓN DE CARACTERÍSTICAS

## BUENAS REPRESENTACIONES



- |                            |   |  |
|----------------------------|---|--|
| Conocimiento del Problema: | → | Buenas Características                 |
|                            | → | Clases Separables (o Casi Separables)  |
| Falta de Conocimiento:     | → | Malas Características (Muchas, además) |
|                            | → | Clases difícilmente separables         |



## 🚩 EL PROBLEMA DE LA DIMENSIONALIDAD

- 🌀 Clasificación de gatos y perros con 10 instancias ... usando un clasificador lineal

- 🌀 Podemos empezar por una sola característica, por ejemplo, el color promedio de 'rojo' en la imagen:



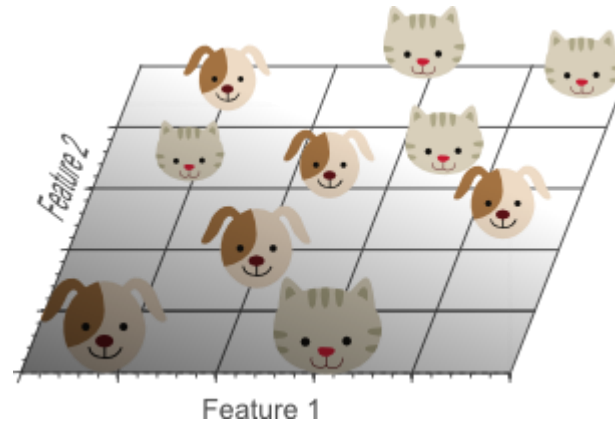
Una sola característica no resulta en una separación perfecta de nuestros datos de entrenamiento.

Tomado de Computer vision for dummies: <http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/>

## 🚩 EL PROBLEMA DE LA DIMENSIONALIDAD

🌀 Clasificación de gatos y perros con 10 instancias ...

🌀 Decidimos añadir otra característica, por ejemplo, el color promedio de "verde" en la imagen:

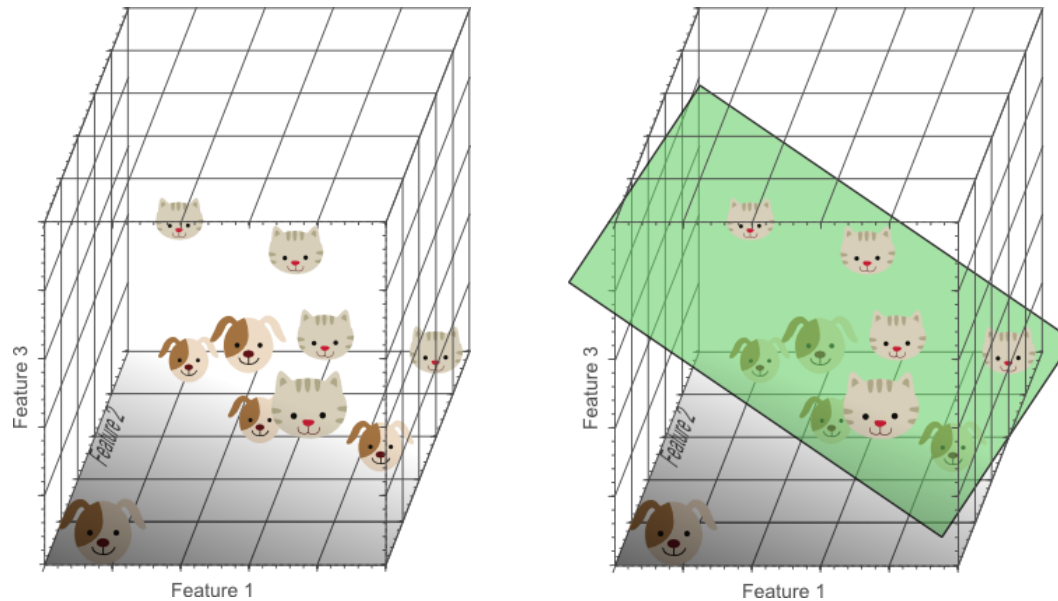


Aún agregando una segunda característica **el problema NO es linealmente separables**: No hay una sola línea que pueda separar todos los gatos de todos los perros en este ejemplo.

## 🦋 EL PROBLEMA DE LA DIMENSIONALIDAD

🌀 Clasificación de gatos y perros con 10 instancias ...

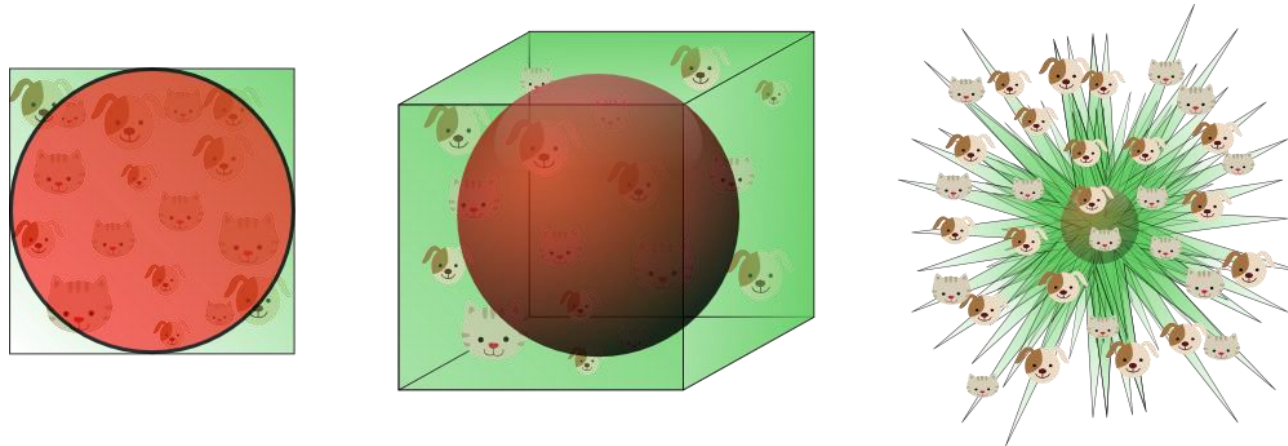
🌀 Finalmente, decidimos añadir una característica más: el color promedio de "azul" en la imagen:



Agregar una tercera característica resulta en un problema separable: existe un plano que separa perfectamente perros de gatos

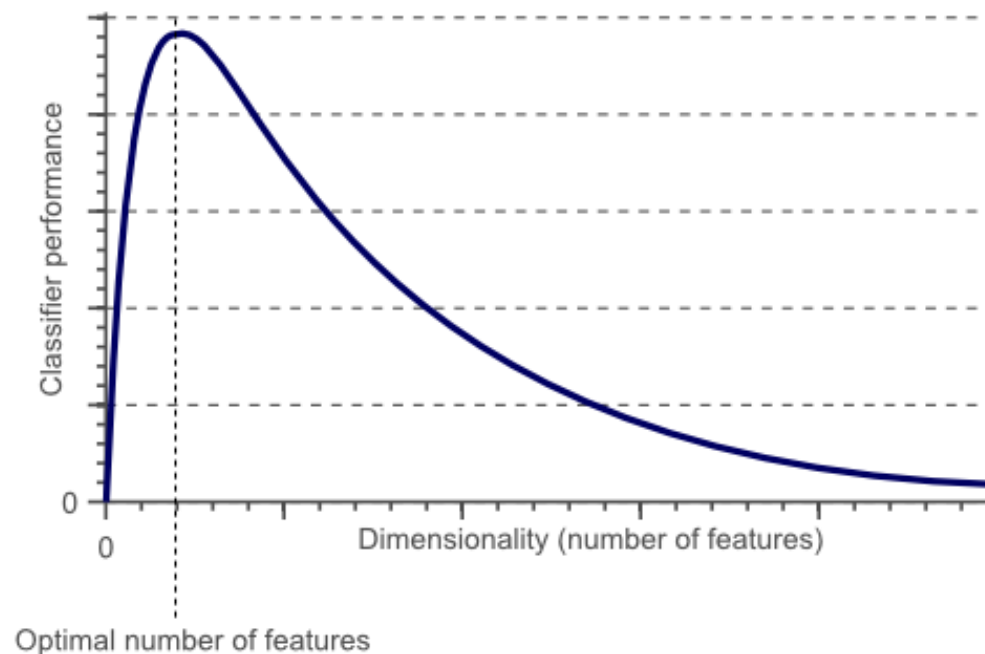
## 🚩 EL PROBLEMA DE LA DIMENSIONALIDAD

- 🌀 Clasificación de gatos y perros con 10 instancias ...
- 🌀 Sin embargo, tenga en cuenta cómo la densidad de las muestras de entrenamiento disminuye exponencialmente cuando aumentamos la dimensionalidad del problema.



Como la dimensionalidad aumenta, un mayor porcentaje de los datos se ubica en las esquinas del espacio: se dispersan los datos

## 🏆 EL PROBLEMA DE LA DIMENSIONALIDAD



La **MALDICIÓN DE LA DIMENSIONALIDAD**: cuando aumenta la dimensionalidad, el volumen del espacio aumenta exponencialmente haciendo que los datos disponibles se vuelven dispersos





## ¿POR QUÉ SELECCIONAR CARACTERÍSTICAS?

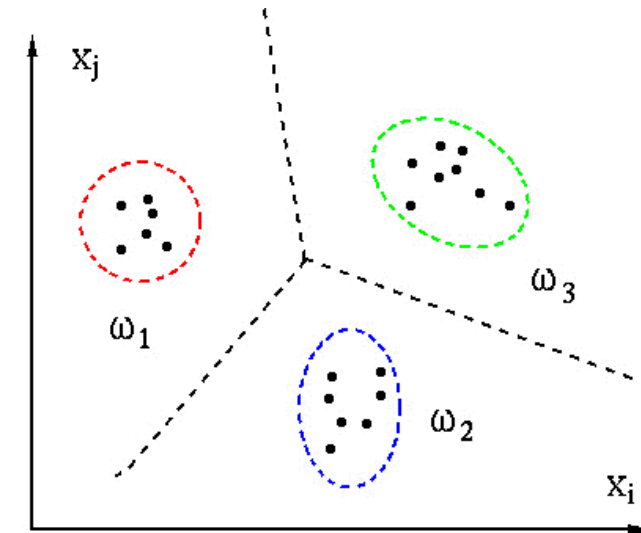
## 🚀 EL PORQUE ...

🌀 ¿Por qué reducir la dimensionalidad del problema?

- 🌀 **Precisión** → Un conjunto pequeño de características define un espacio menor
- 🌀 **Velocidad** → Menos características exigen menos mediciones
- 🌀 **Memoria** → Menos características menos memoria para su almacenamiento

🌀 ¿Cómo reducir el número de características utilizado?

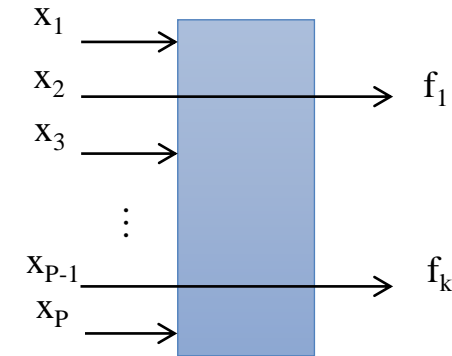
- 🌀 Eliminar características “ruidosas”
- 🌀 Medición del costo



## 🏆 APROXIMACIONES A LA REDUCCIÓN DE LA DIMENSIONALIDAD

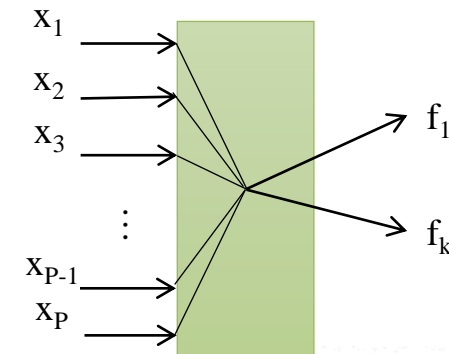
### 🎯 Selección de Características: **Selecione** k características de P:

- 🌀 Ventajas
  - ✓ Disminuye las mediciones
  - ✓ Fácil interpretación
- 🌀 Desventajas
  - ✓ Métodos costosos computacionalmente
  - ✓ Son aproximaciones



### 🎯 Extracción de Características: **Transforme** P características en k:

- 🌀 Ventajas
  - ✓ Suelen ser menos costosas
  - ✓ Pueden ser No-Lineales
- 🌀 Desventajas
  - ✓ Necesita todas las características
  - ✓ Sub-óptimas dependientes de un criterio



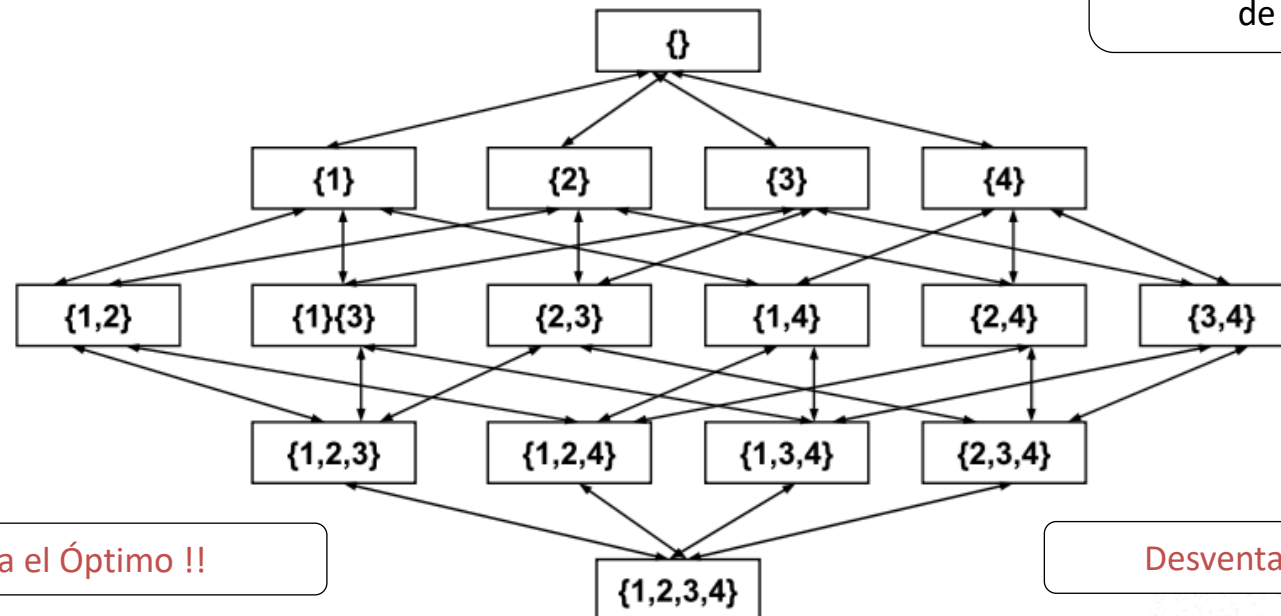


# MÉTODOS DE SELECCIÓN

## MÉTODOS DE SELECCIÓN

- Estos métodos seleccionan un **subconjunto** de  $k$  de características optimizando un criterio determinado (o función objetivo).
- La solución trivial es **BÚSQUEDA EXHAUSTIVA**: evalúa **TODOS** los posibles subconjuntos de características. ¿Cuántos hay ? ...

$$n_k = \frac{P!}{(P-k)!k!} \text{ Subconjuntos de tamaño } k$$



Garantiza el Óptimo !!

Desventaja: Complejidad

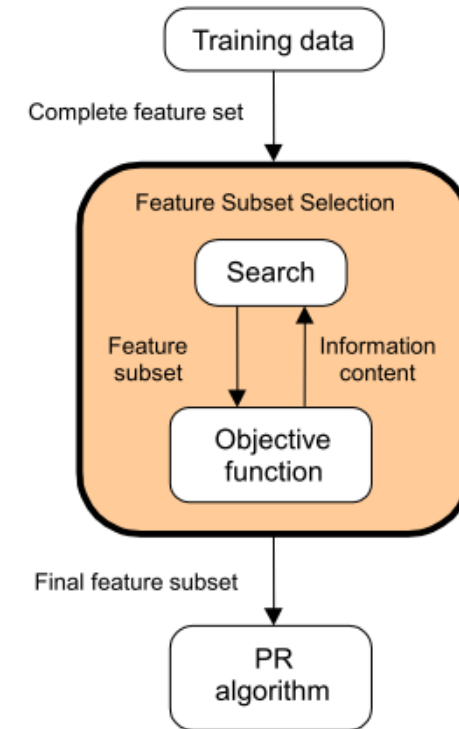
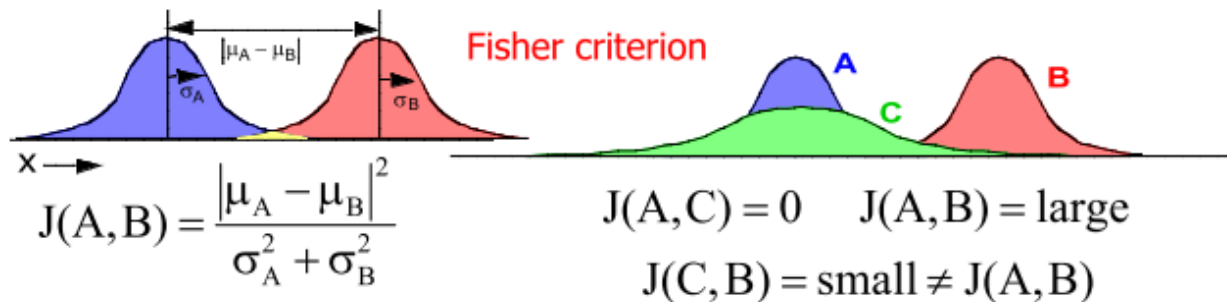


## MÉTODOS DE SELECCIÓN

**FUNCIÓN OBJETIVO:** evalúa subconjuntos candidatos y devuelve una medida de su "bondad".

Esta retroalimentación se utiliza por el estrategia de búsqueda para seleccionar nuevo candidatos.

Una función simple: El criterio de Fisher

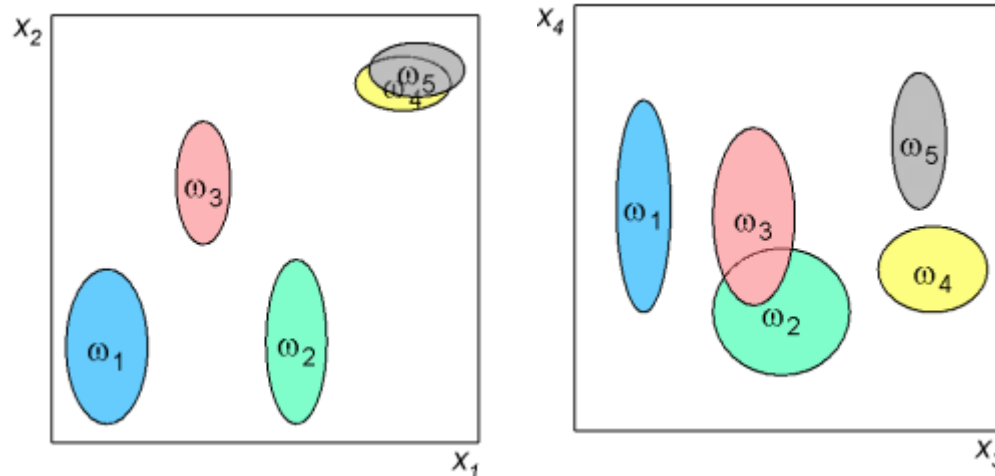


## ✦ MÉTODOS DE SELECCIÓN

🌀 **SELECCIÓN SIMPLE:** Seleccionar las mejores  $k$  características individuales.

🌀 Ventaja: Rápido de Calcular

🌀 Desventaja: Combinando variables que individualmente son buenas, no siempre conduce a buenos resultados

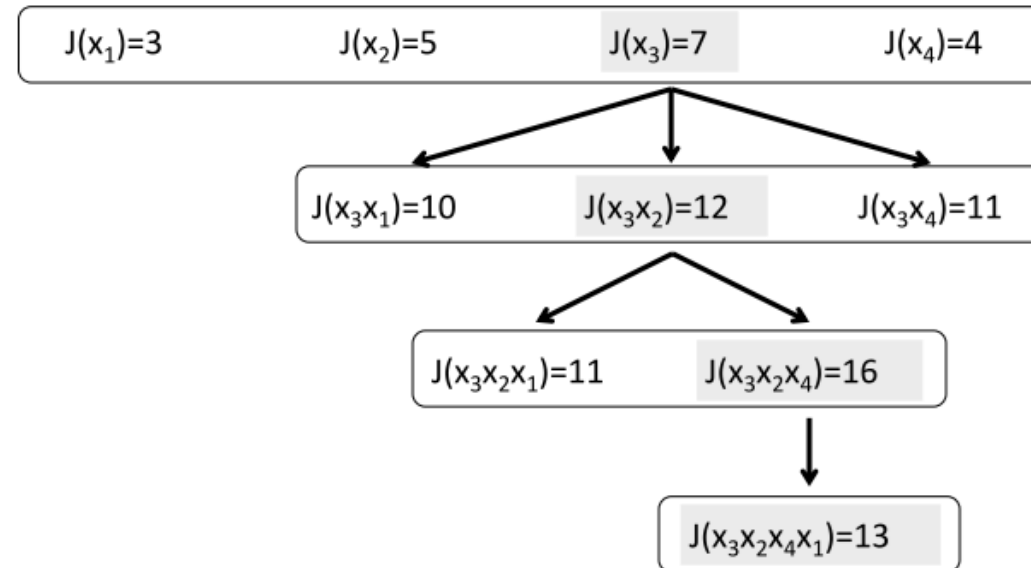


- Cualquier función buena haría un ranking como este:  $J(x_1) > J(x_2) \cong J(x_3) > J(x_4)$
- Solución óptima:  $\{x_1, x_4\}$
- Solución del algoritmo:  $\{x_1, x_2\}$

## MÉTODOS DE SELECCIÓN

### SELECCIÓN HACIA ADELANTE (FS):

- Comience con un conjunto vacío.
- Adicione una característica a la vez tal que se maximice  $J$

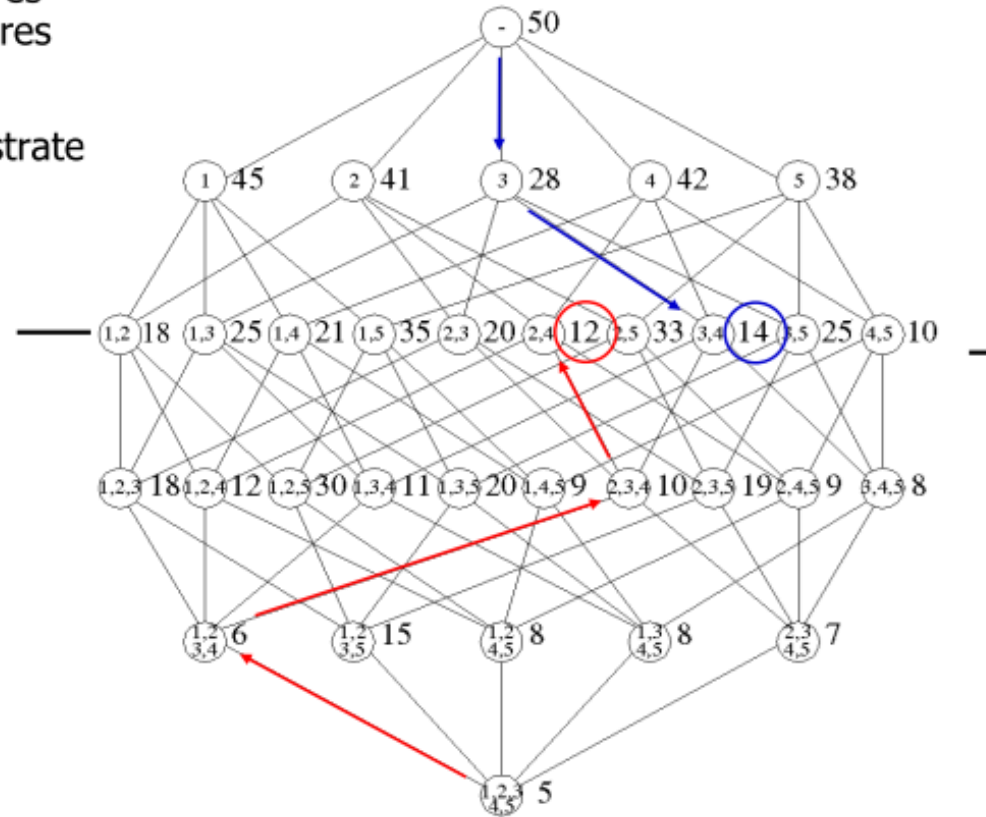


### SELECCIÓN HACIA ATRÁS (BS):




- Comience el conjunto de todas las características
- Elimine una característica a la vez tal que se maximice  $J$  al eliminarla

## ✈ MÉTODOS DE SELECCIÓN

- Select  $l = 2$  features out of  $k = 5$  features
- Suboptimality illustrate
  - forward
  - backward



## MÉTODOS DE SELECCIÓN




-  **SELECCIÓN BRANCH & BOUND:** es una búsqueda hacia atrás con retroceso que garantiza encontrar el subconjunto óptimo de característica bajo el supuesto de monotonidad.
-  Inicia con el conjunto completo de características formado un árbol de evaluación.
-  Los nodos cuya función objetivo son menores que la mejor solución actual no se exploran por el supuesto de monotonidad el cual asegura de que los hijos de esos nodos no van a contener una solución mejor.

Ver ejemplo de Domingo Mery



## MÉTODOS DE SELECCIÓN

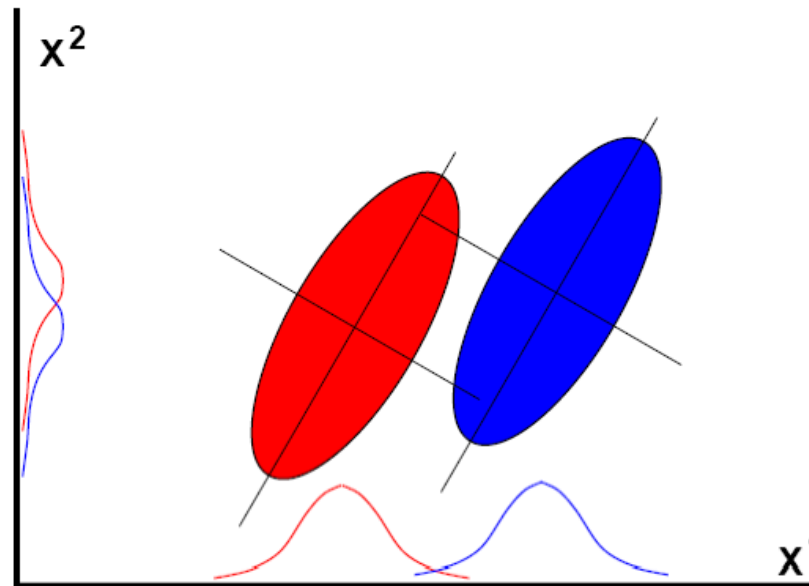
### EN RESUMEN:

-  El número de características normalmente se limita en el aprendizaje debido a los costos y a la precisión del clasificador (maldición de la dimensionalidad)
-  Encontrar buenas características **no es trivial**. La selección de características puede mejorar el rendimiento y ayudar a la interpretación:
  - **Requisitos:** criterio y el algoritmo de búsqueda.
  - **Estimación de rendimiento:** por ejemplo, validación cruzada.
-  **Pasos:**
  - Conocimiento especializado (por ejemplo, iniciar con 100 caractrísticas)
  - Ranking Individual (Seleccionar 40)
  - Selección hacia adelante (Seleccionar 20)
  - Selección hacia atrás usando branch & bound (10)

## MÉTODOS DE EXTRACCIÓN: PCA

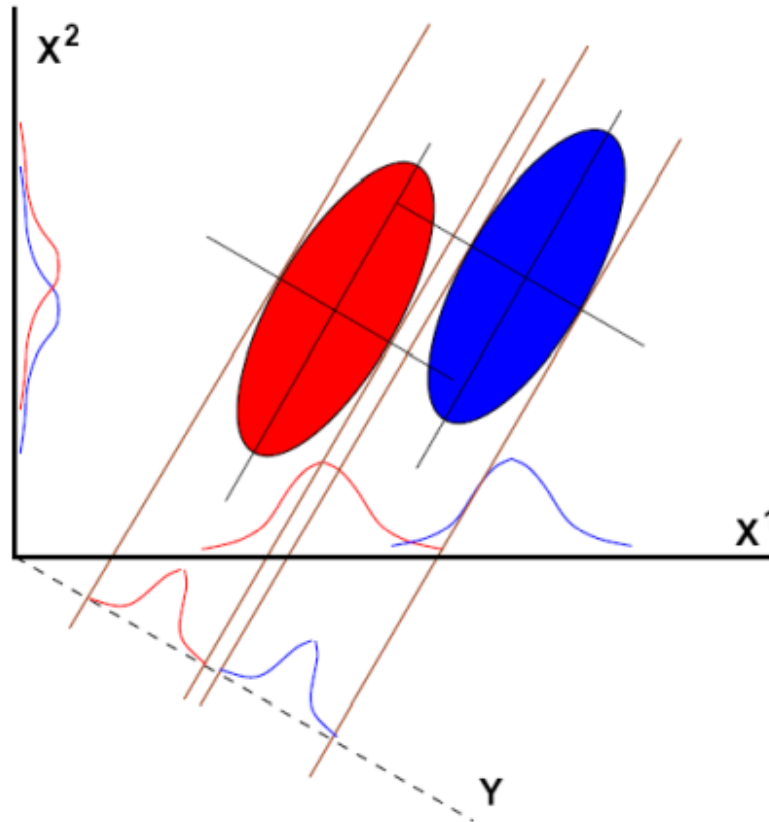
## 🚀 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

- 🌀 Hay casos en los que la selección de cualquiera de las variables originales no proporciona una buena solución.



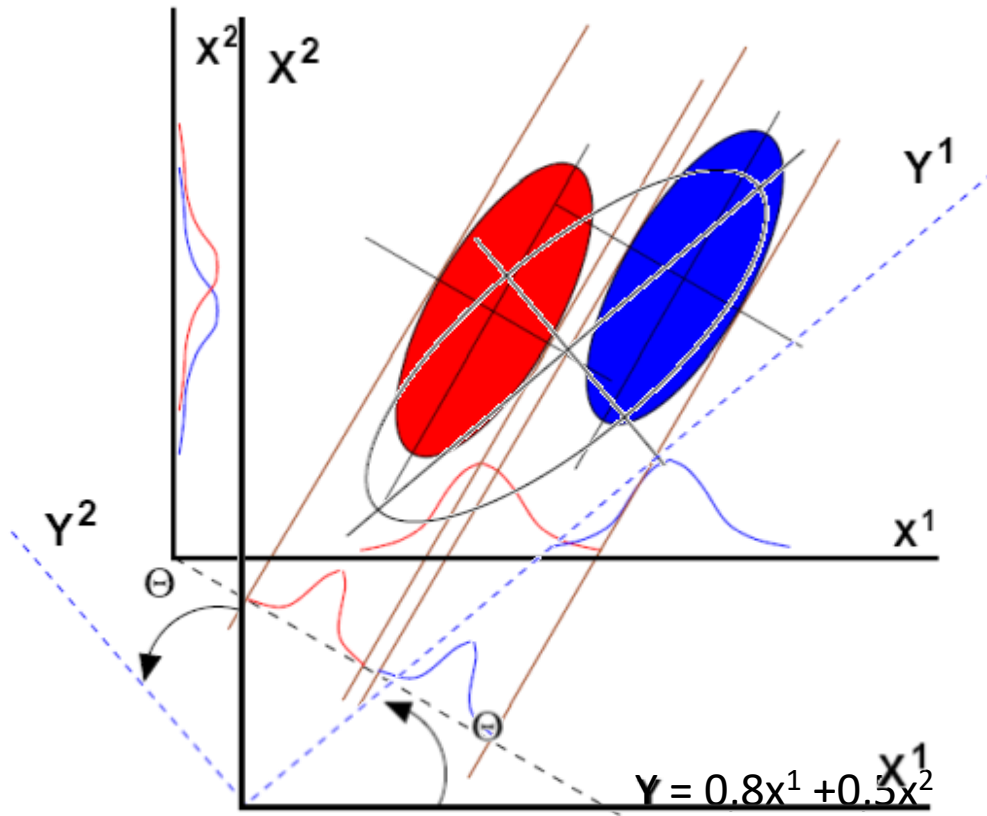
## 🚀 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

- 🕒 Si transformamos el espacio de manera lineal usando una nueva variable  $Y$  que tenga la forma  $Y = 0,8x^1 + 0,5x^2$ , se obtienen mejores resultados en la separación:



## ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

- PCA es el método más popular de extracción de características
- PCA es una transformación lineal



**Objetivo:** Transformar el espacio de representación  $P$  en uno nuevo  $P'$ , en el que los datos estén incorrelacionados, es decir, que su matriz de sea diagonal.



Es decir, se trata de encontrar un nuevo conjunto de ejes ortogonales en el que la varianza de los datos sea la máxima en alguna dirección.

## 🚀 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

🌀 La relación entre los nuevos ejes y los antiguos consiste en:

$$Y_1 = W_{11}X_1 + W_{12}X_2$$
$$Y_2 = W_{21}X_1 + W_{22}X_2$$

- 🌀 La mayor parte de la información contenida en el espacio  $P$  puede retenerse únicamente en el primer eje principal  $Y_1$ , lo que implica una selección de características ...
- 🌀 Los nuevos ejes se definen secuencialmente de manera que un nuevo eje se define como aquel que es perpendicular al seleccionado anteriormente y su dirección es la de la máxima varianza de entre todos los ejes posibles.

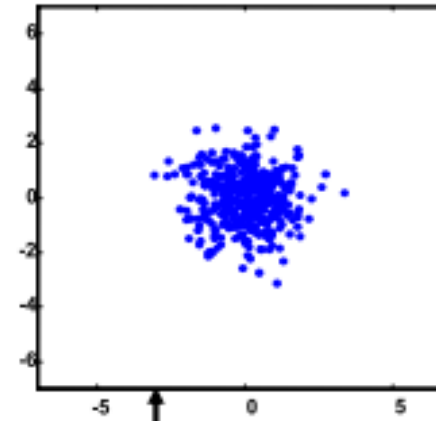
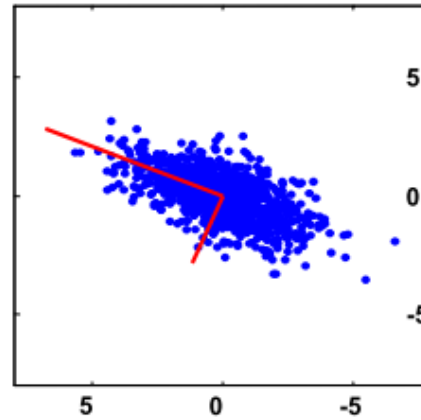
<http://www.visiondummy.com/2014/03/eigenvalues-eigenvectors/>

## ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Encontrar  $W_{ij}$ :

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

- Los ejes que definen  $P'$  son ortogonales
- Los datos en  $P'$  están incorrelacionados
- Se debe minimizar el error de reconstrucción



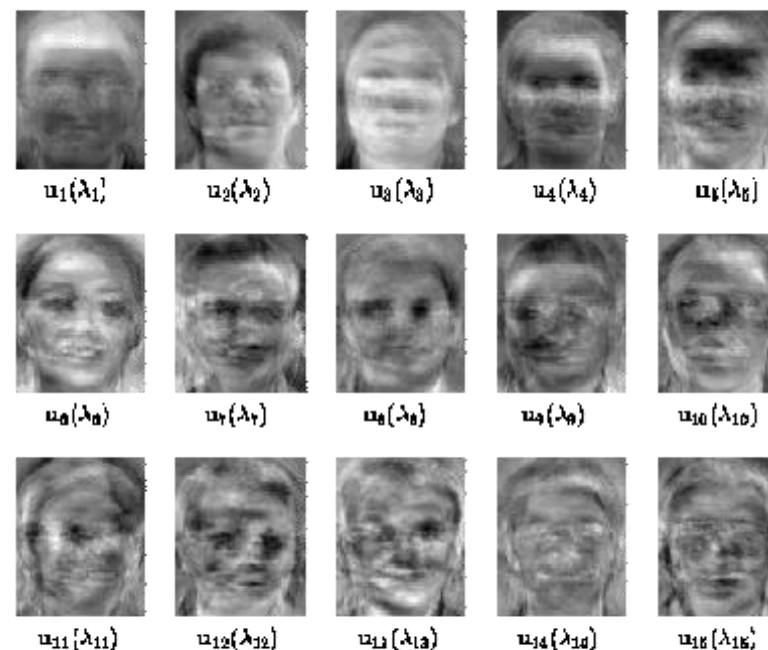
- Los **componentes principales** son los primero  $k$  vectores propios de la transformación ...



## ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Ejemplo con Imágenes para Reconocimiento de un Patrón en las imágenes

*Eigenfaces for Recognition*, Turk, M. & Pentland, A. ,  
*Journal of Cognitive Neuroscience*, 3, 71-86, 1991.





Institución  
**Universitaria**  
Reacreditada en Alta Calidad

¡Gracias!

Hacia una era de  
**Universidad y**  
*Humanidad*



Alcaldía de Medellín