

# Clasificación eficiente de objetos usando la Característica de Euler

**Erik Amézquita**<sup>1</sup>   Mario Canul<sup>2</sup>   Antonio Rieser<sup>3</sup>  
`erik.amezquita@cimat.mx`

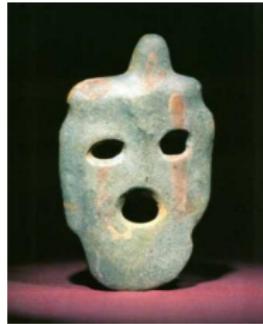
<sup>1</sup>DEMAT, UGto

<sup>2</sup>CIMAT

<sup>3</sup>CONACYT-CIMAT

XVI EPE, 13 de abril 2018

# Pregunta, Problema y Objetivo



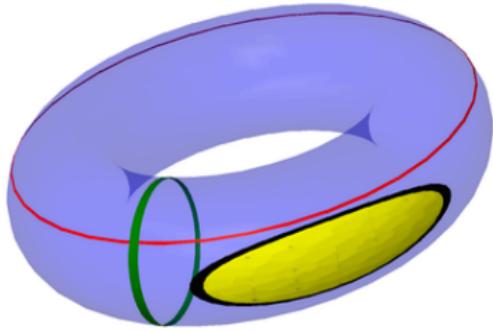
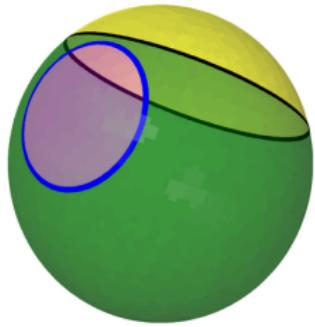
- Entre 1978 y 1982 se halló una ofrenda de 162 máscaras en el Templo Mayor (CDMX).
- No es claro cuántas culturas están representadas ni cuáles.
- La clasificación actual es subjetiva.
- ¿Puede la topología decirnos algo de estas máscaras?
- **Sí, con Análisis Topológico de Datos (ATD)**

# Vistazo general

- 1 Introducción
- 2 Homología Simplicial
- 3 Gráfica de Característica de Euler
- 4 Máquinas de Soporte Vectorial (SVM)
- 5 Caso arqueológico
- 6 Conclusiones

# Homología Simplicial

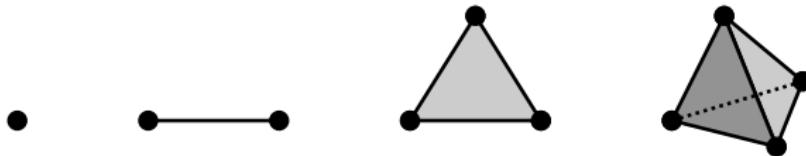
- Nos interesa caracterizar los agujeros de alguna manera.
- Ver que los ciclos en la esfera y toro son distintos.



# Simplejos y Complejos simpliciales

- Vértices  $\mathbf{v}_0, \dots, \mathbf{v}_d \in \mathbb{R}^d$  en posición general.
- El  **$d$ -simplejo** es la envolvente convexa de los vértices.

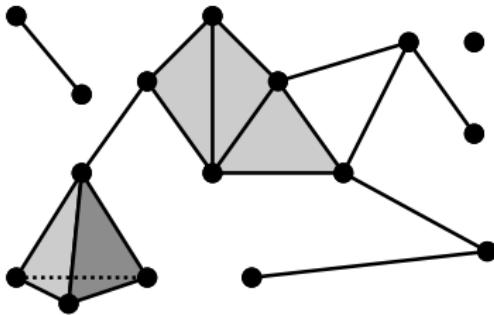
$$S_d := \left\{ \sum_{i=0}^d \lambda_i \mathbf{v}_i : \lambda_i \geq 0, \sum_{i=0}^d \lambda_i = 1 \right\} \quad (2.1)$$



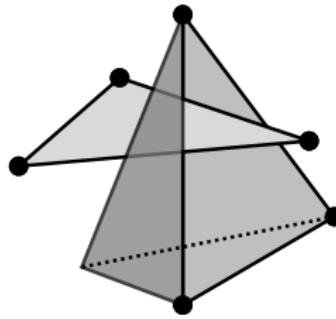
$d$ -simplejos para  $d = 0, 1, 2, 3$

- $\sigma = (\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_d)$ .
- $\tau = (\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{d-1})$  es una **cara** de  $\sigma$ .

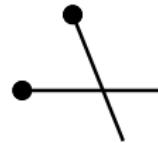
- Un **complejo simplicial** es un conjunto de simplejos bien pegados.
  - Están todas las caras de todo simplejo.
  - Dos simplejos adyacentes comparten cara completa.



(a) Bien

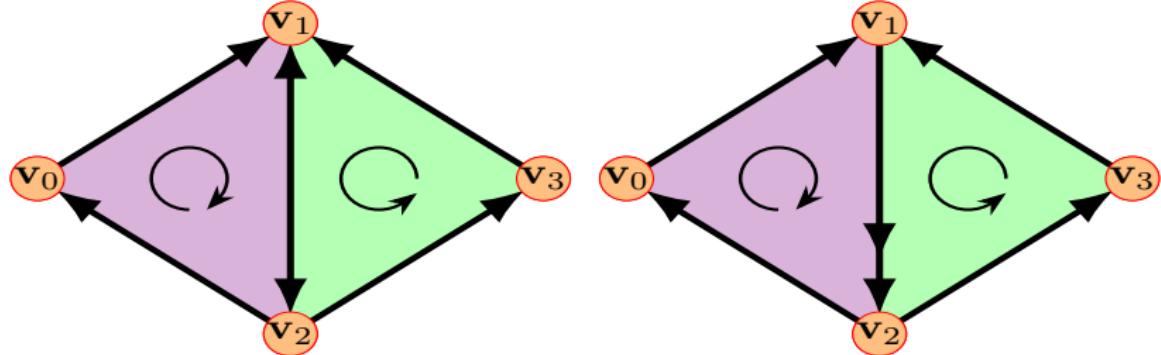
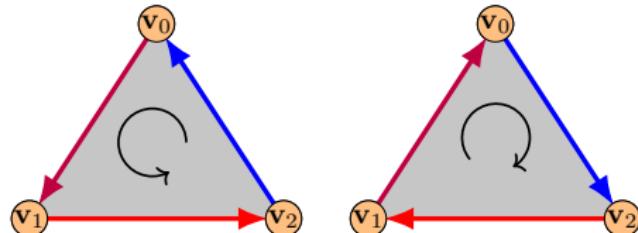


(b) Mal



# Orientación y Compatibilidad

- $\sigma = (\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_d)$  indica también **una de dos posibles** orientaciones.

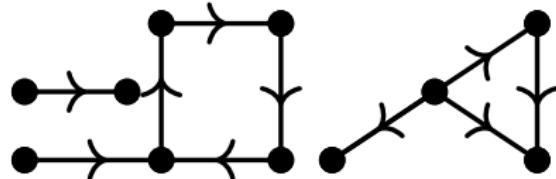


(a) Triángulos orientados compatibles

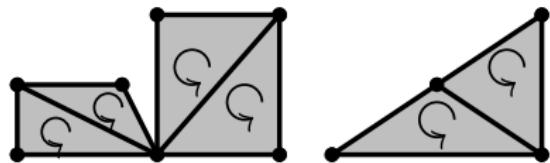
(b) Incompatibilidad

# Cadenas y Homomorfismos

- Grupo de  **$q$ -cadenas**  $C_q = \left\{ \sum_{i=1}^k \lambda_i \sigma_i : \lambda_i \in \mathbb{Z}, \sigma_i$  un  $q$ -simplejo  $\right\}$ .
- $-\sigma$  orientación **opuesta** a  $\sigma$ .
- Definimos **homomorfismos**  $\varphi$  en los simplejos y extendemos **linealmente** al resto de cadenas.
- Verificar que  $\varphi(-\sigma) = -\varphi(\sigma)$ .
- $\varphi(A) = \sum \lambda_i \varphi(\sigma_i)$



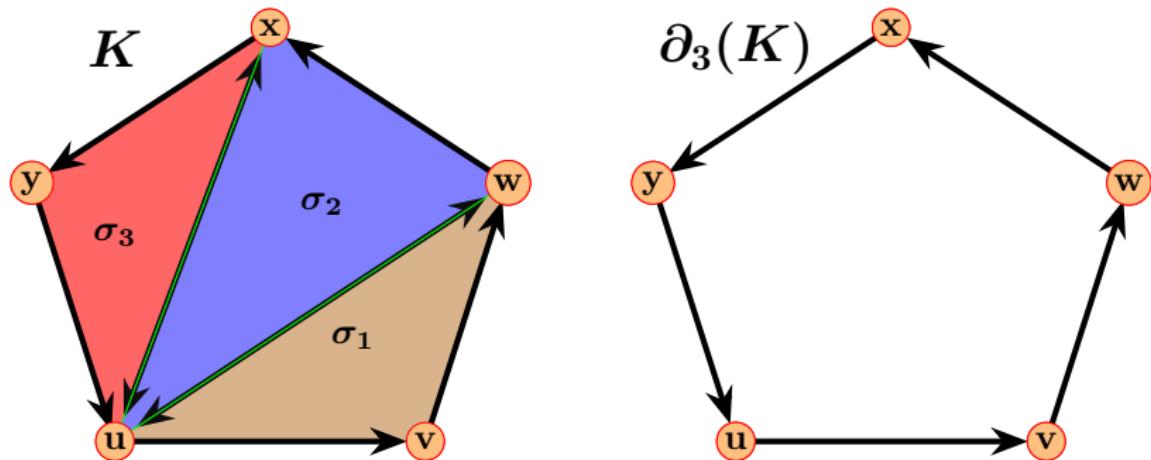
(a) 1-cadena



(b) 2-cadena

# Homomorfismo frontera $\partial_q : C_q \rightarrow C_{q-1}$

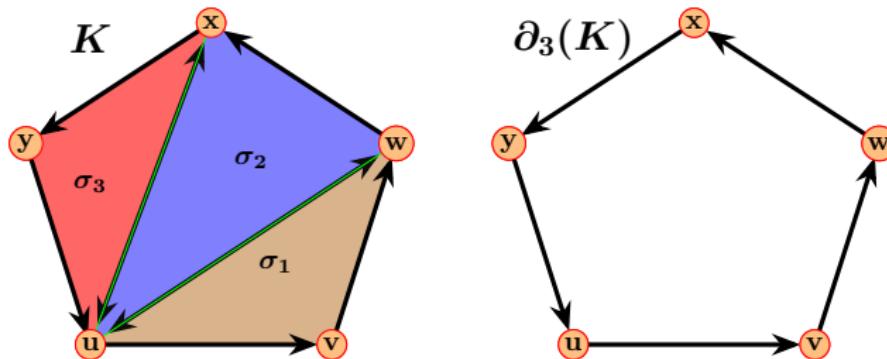
- $\partial\sigma = \sum_{i=0}^q (-1)^i (\mathbf{v}_0, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_n).$
- $\partial^2 = \partial_q \circ \partial_{q-1} = 0.$



$$\partial_2(\partial_3(K)) = (\mathbf{v} - \mathbf{u}) + (\mathbf{w} - \mathbf{v}) + (\mathbf{x} - \mathbf{w}) + (\mathbf{y} - \mathbf{x}) + (\mathbf{u} - \mathbf{y}) = \mathbf{0}.$$

# Ciclos, Fronteras y Homología

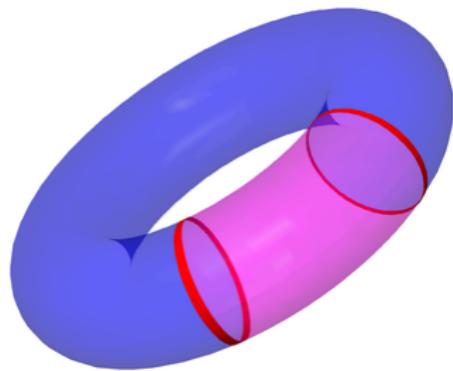
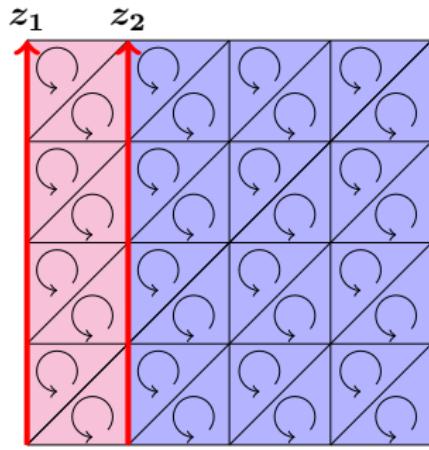
- $Z_q(K) = \ker \partial_q$  es el grupo de  **$q$ -ciclos**
- $B_q(K) = \text{im } \partial_{q+1}$  es el grupo de  **$q$ -ciclos frontera**
- $H_q(K) = Z_q(K)/B_q(K)$  es el  $q$ -ésimo grupo de **homología**.
- $H_q(K) \simeq F \oplus T$  por ser grupo abeliano finitamente generado.
- $\beta_q(K) = \dim(H_q(K))$  es el  $q$ -ésimo **número de Betti**.



$$\partial_2(\partial_3(K)) = (v - u) + (w - v) + (x - w) + (y - x) + (u - y) = \mathbf{0}.$$

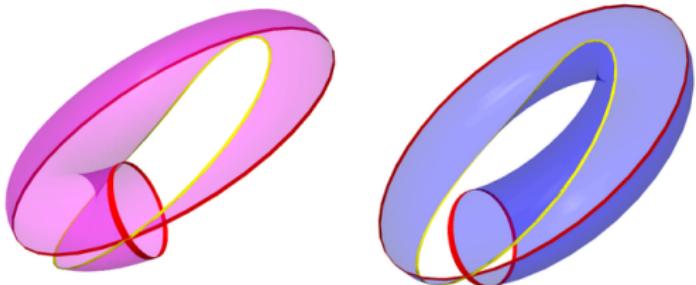
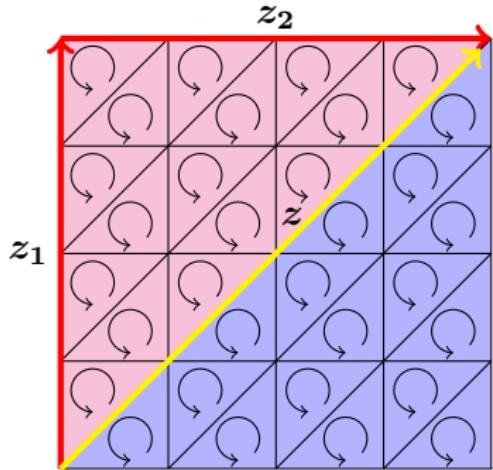
$$H_q(K) = Z_q(K)/B_q(K)$$

- Dos ciclos son **homológicos** si su diferencia es un ciclo frontera.



Los ciclos rojos son homológicos en el toro

$$H_1(T) = \mathbb{Z} \oplus \mathbb{Z} \quad \text{y} \quad \beta_1(T) = 2$$

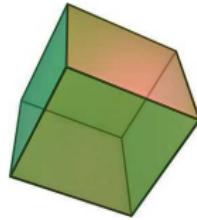


$z_1 + z_2 + z$  es un ciclo frontera definido por la 2-cadena rosa

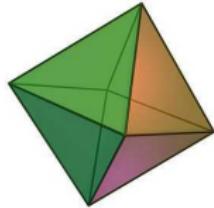
- En general,  $\beta_q$  cuenta el **número de agujeros** homológicamente distintos!

# La Característica de Euler

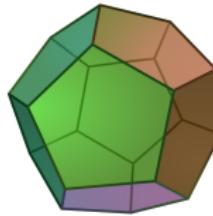
- Supongamos que  $K = \bigcup_{q=0}^d V_q(K)$ .
- $\chi(K) = \sum_{q=0}^d (-1)^q |V_q(K)|$  es la **característica de Euler**.
- $\chi(K) = \sum_{q=0}^d (-1)^q \beta_q(K)$ .



(a)  $8 - 12 + 6 = 2$



(b)  $6 - 12 + 8 = 2$



(c)  
 $20 - 30 + 12 = 2$



$$\chi(S^2) = 1 - 0 + 1 = 2$$

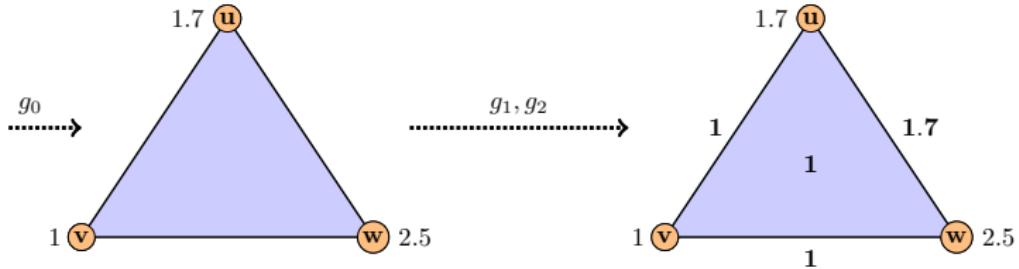
# Gráfica de Característica de Euler: Filtraciones

- Idea propuesta por Richardson y Weirman en [4]
- Fijamos una función de filtración  $g$  para los vértices y luego la extendemos al resto de los  $q$ -simplejos:

$$g_q(\{v_0, v_1, \dots, v_q\}) = \min_{0 \leq i \leq k} \{ g(v_i) \}$$

$g_q : V_q \rightarrow [a, b]$        $q$ -simplejo

Una función  $g : V_0 \rightarrow [a, b]$   
 $V_0$  el conjunto de vértices;  $[a, b]$  intervalo fijo.



# GCE: Umbrales

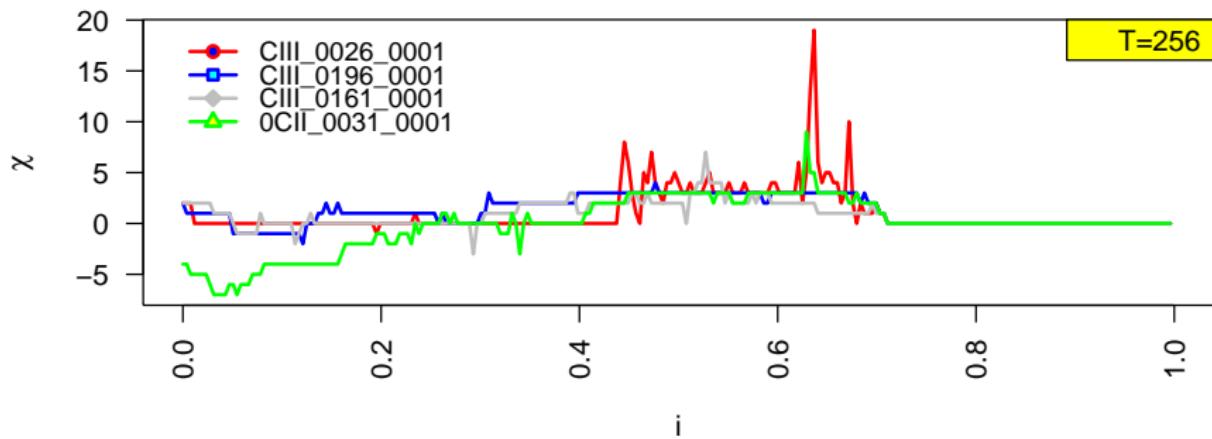
- $T$  umbrales equiespaciados  $a = t_0 < t_1 < \dots < t_T = b$ .
- Para cada  $t_i$  definimos  $V_q^{(i)} = \{\sigma \in V_q : g(\sigma) > t_i\}$ .
- $\chi^{(i)} = \sum_{q=0}^d (-1)^q |V_q^{(i)}|$  es la **Característica de Euler al  $i$ -ésimo umbral**.



Filtro de  $1/(10\text{mo Vecino Más Cercano})$  (KNN) con gorila

- La **Gráfica de Característica de Euler** es la gráfica  $\chi_i$  vs  $t_i$ .

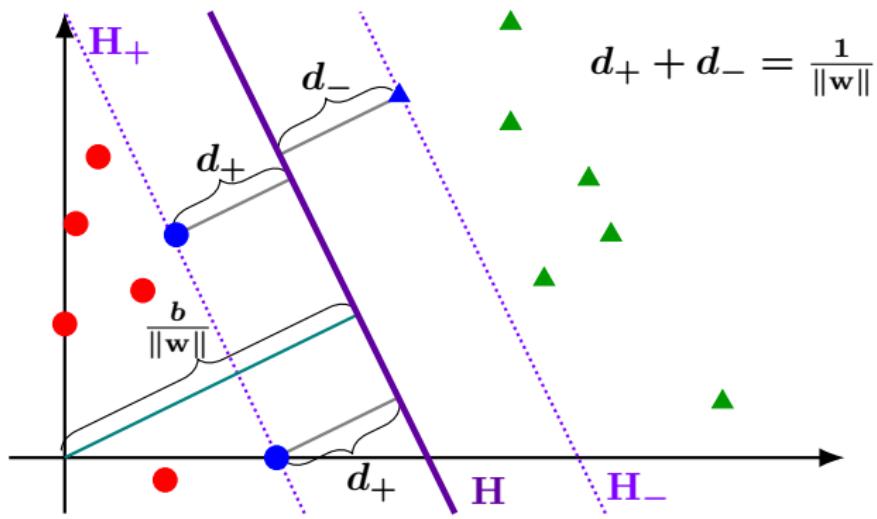
$$\text{ECG: } 1 - (x^2 + y^2 + z^2)$$



- Cada GCE puede pensarse como el **vector**  $(\chi_0, \chi_1, \dots, \chi_{T-1}) \in \mathbb{R}^T$ .

# SVM: Caso lineal separable

- Tenemos  $n$  puntos etiquetados  $\{\mathbf{x}_i, y_i\}_{i=1}^n \subset \mathbb{R}^d$  con  $y_i \in \{-1, +1\}$ .
- Debemos hallar el **hiperplano  $H$**  que los separa mejor.
- $H$  está definido por un vector normal  $\mathbf{w}$  y un escalar  $b$ .
- $H = \{\mathbf{x} : \langle \mathbf{x}, \mathbf{w} \rangle + b = 0\}$ .



# Optimización restringida

$$\min_{(\mathbf{w}, b) \in \mathbb{R}^d \times \mathbb{R}} f(\mathbf{w}, b) := \frac{1}{2} \|\mathbf{w}\|^2,$$

tal que  $c_i(\mathbf{w}, b) := y_i(\langle \mathbf{x}_i, \mathbf{w} \rangle + b) \geq 1$  para todo  $i = 1, \dots, n$ .

- Usamos condiciones de **Karush-Kuhn-Tucker** y el **dual de Wolfe**.

$$\max_{\alpha_i \geq 0} \sum_{1 \leq i \leq n} \alpha_i - \frac{1}{2} \sum_{1 \leq i, j \leq n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle,$$

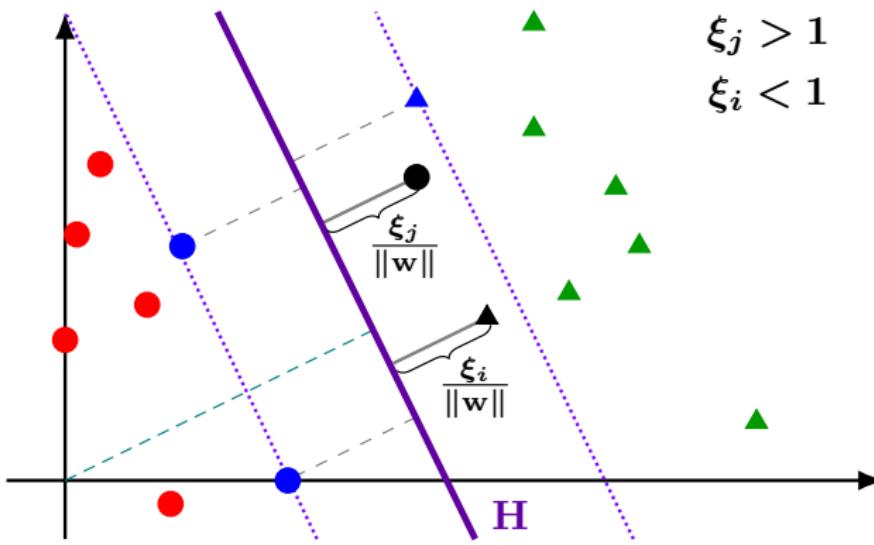
donde

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i.$$

$$b = y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle \quad \text{para } \alpha_i > 0.$$

## SVM: Caso lineal no separable

- Ahora consideramos errores  $\xi_i \geq 0$  para cada punto.
- Hay error cuando  $\xi_i > 1$ .



# Mismas técnicas de optimización

$$\min_{(\mathbf{w}, b, \xi) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d} f(\mathbf{w}, b, \xi) := \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=0}^n \xi_i, \quad C > 0,$$

tal que  $y_i(\langle \mathbf{x}_i, \mathbf{w} \rangle + b) + \xi_i \geq 1$  para todo  $i = 1, \dots, n$   
y  $\xi_i \geq 0$ .

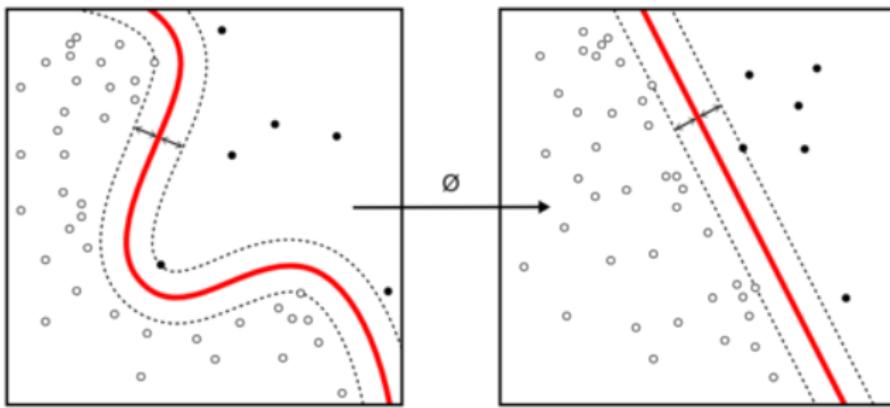
- Con KKT y Wolfe:

$$\max_{0 \leq \alpha_i \leq C} \sum_{1 \leq i \leq n} \alpha_i - \frac{1}{2} \sum_{1 \leq i, j \leq n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle.$$

# SVM: Caso no lineal: kernelización

- $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$ , con  $\mathcal{H}$  espacio de Hilbert de dimensión mayor y ahí resolvemos linealmente.
- Pensamos en una función **kernel**  $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ .

$$K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$$

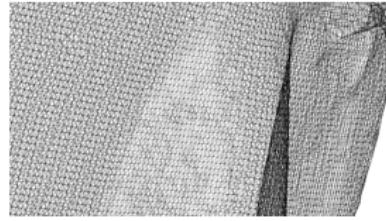
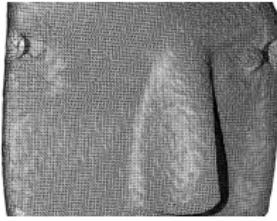


- Teorema de Mercer, Reproducing Kernel Hilbert Spaces (RKHS)

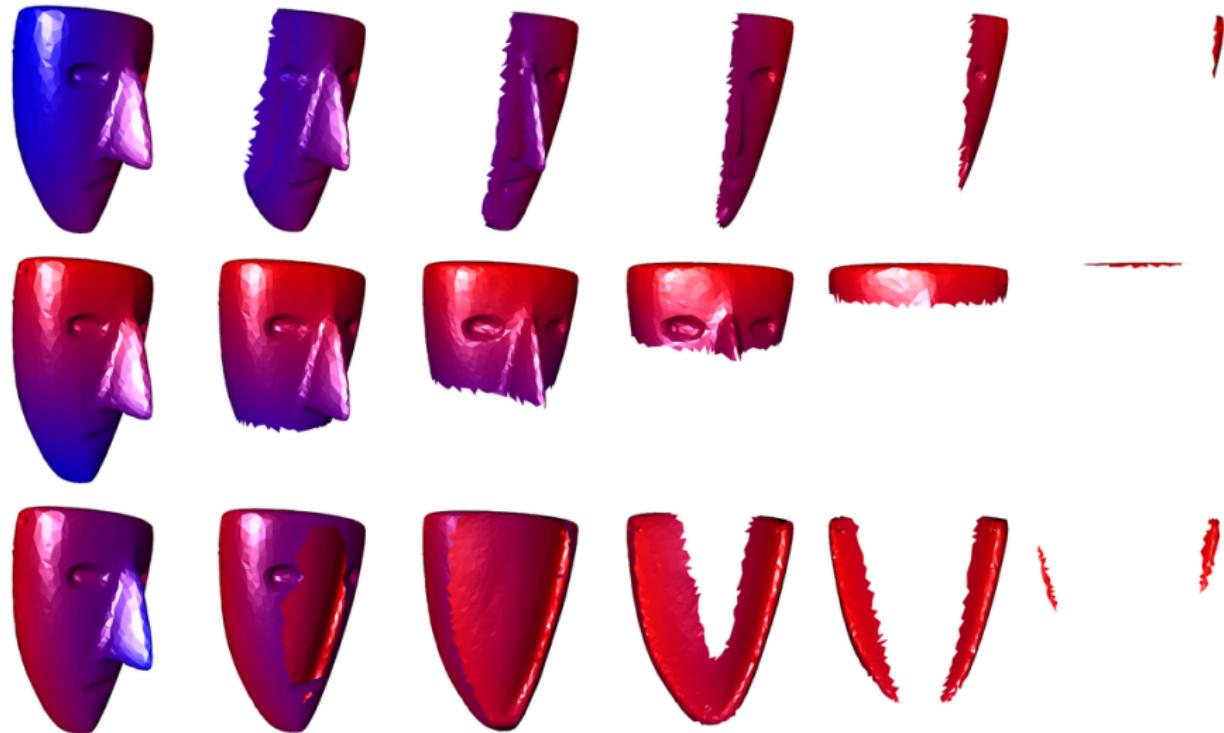
# Ofrendas del Templo Mayor (DF)

- Se cuenta con 128 mallas digitalizadas de máscaras prehispánicas.
- Encajadas en el cubo  $[-1, 1]^3$  con baricentro en el origen.
- 8 familias identificadas y un grupo todavía en discusión.
- Más detalles explicados por D. Jiménez en [3].

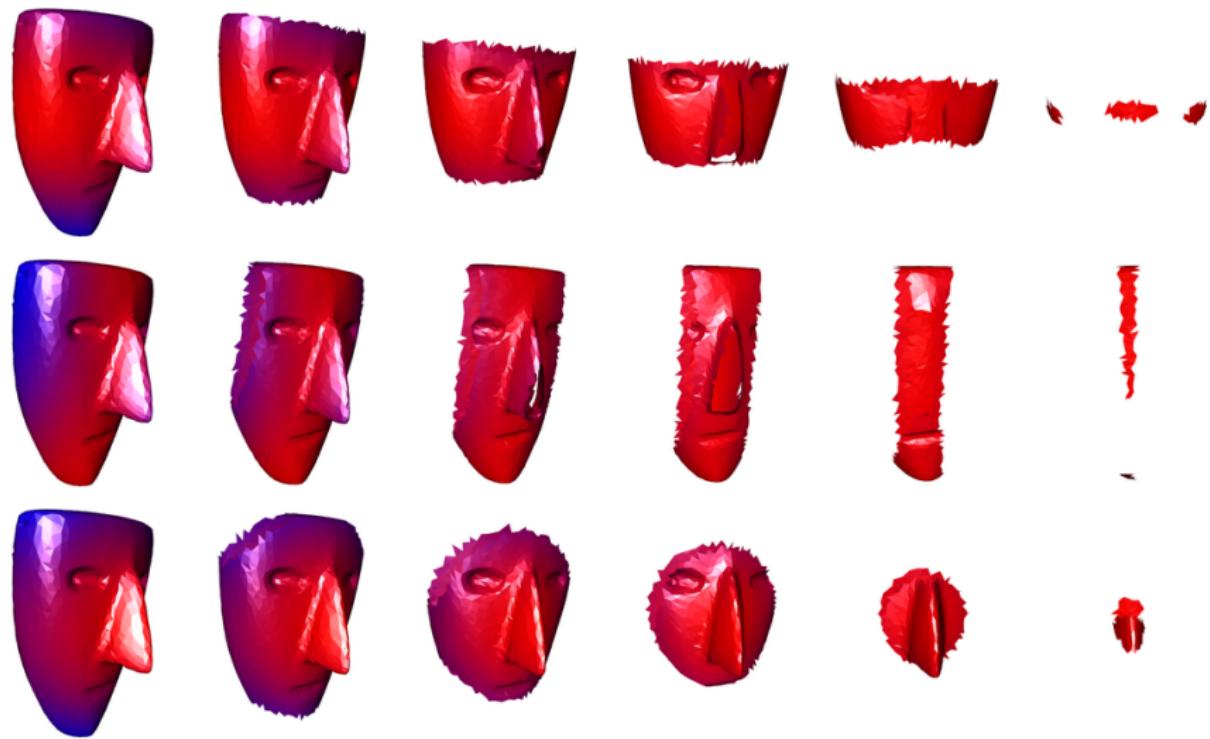
SET	NO. OF ITEMS	SET	NO. OF ITEMS
02	24	07	4
03	6	08	3
04	4	09	7
05	19	10	59
06	2	<b>TOTAL</b>	<b>128</b>



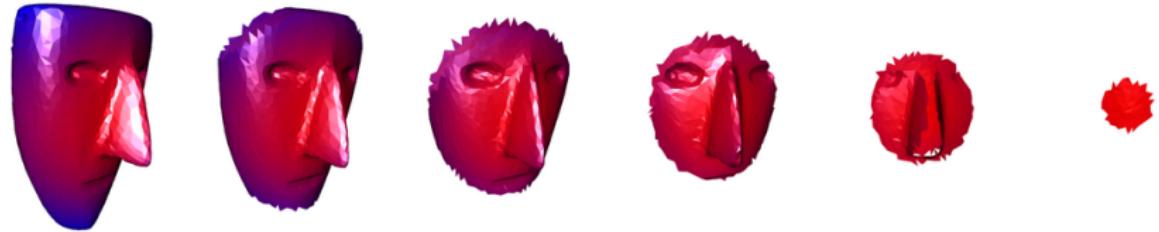
# GCEs: Planos



# GCEs: Cilindros



# GCEs: Esferas



- Se consideraron también filtros basados en valores de curvatura.
- Resultados pésimos.

# Metodología

ENTRENAMIENTO: 8 clases

- Se tomaron parte de las familias 02 y 05.
- Todos los ítems de familias 03, 04, 06, 07, 08.
- Los ítems excluidos son aquellos en que D. Jiménez tiene dudas.

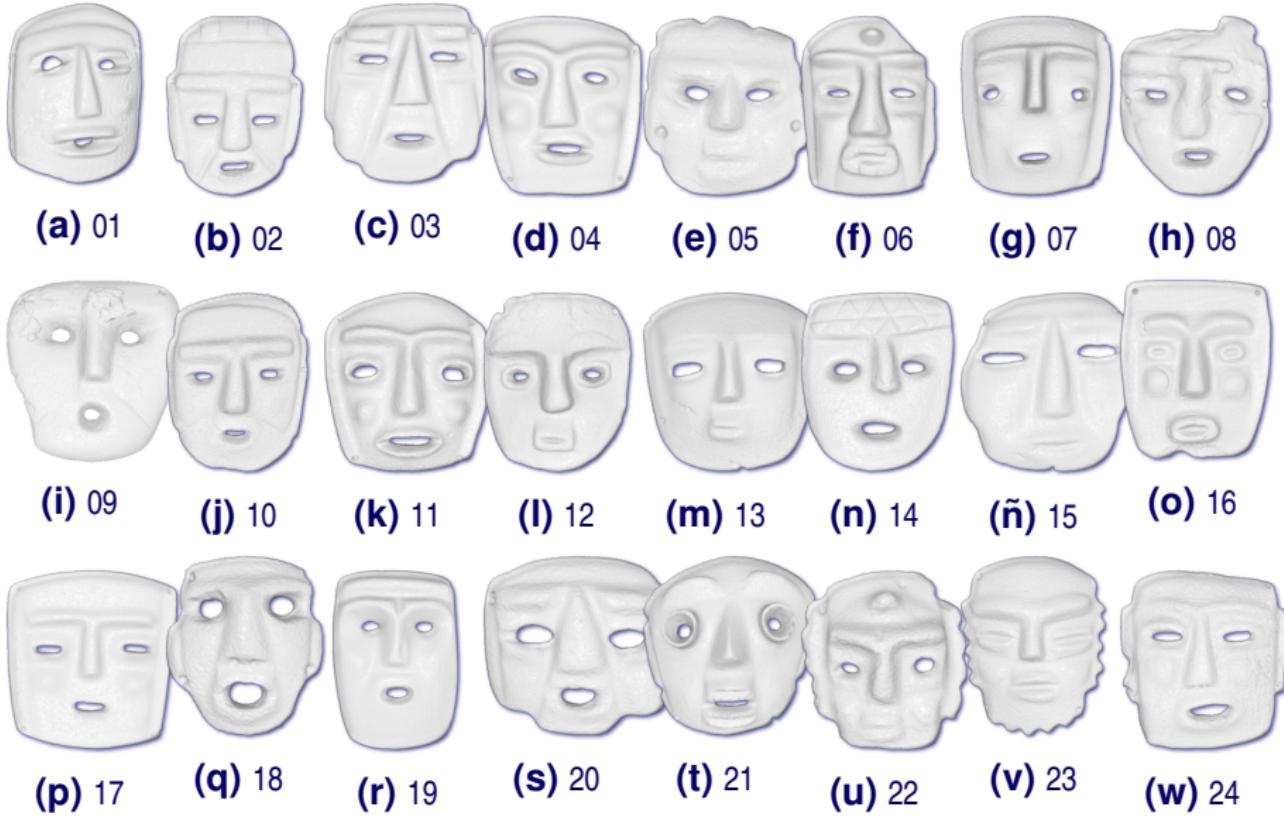
PRUEBA

- Las 128 máscaras en total
- Es de interés como se fracciona la familia 10.

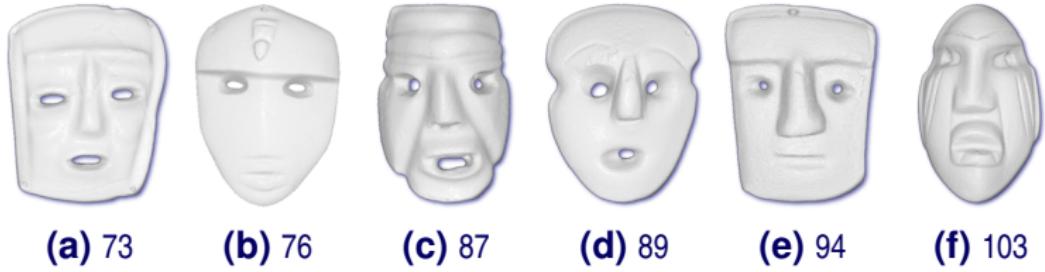
SVMs: 72 evaluaciones distintas

- Kernel polinomial  $(\gamma \langle \mathbf{x}, \mathbf{y} \rangle + k)^{\deg}$ .
- Costo  $C = 10$ .
- Tomar la moda si corresponde al 85 % de respuestas.

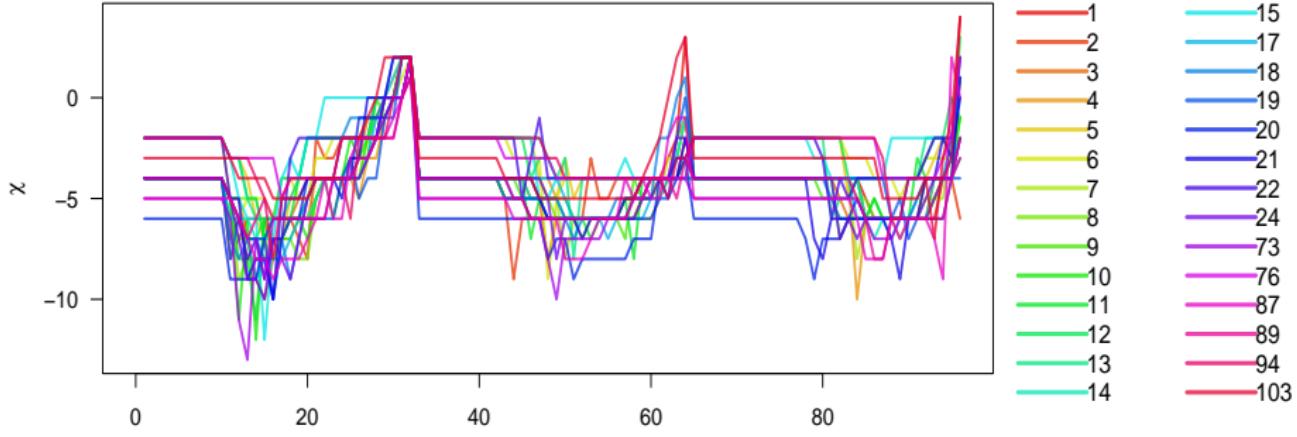
# Familia 02



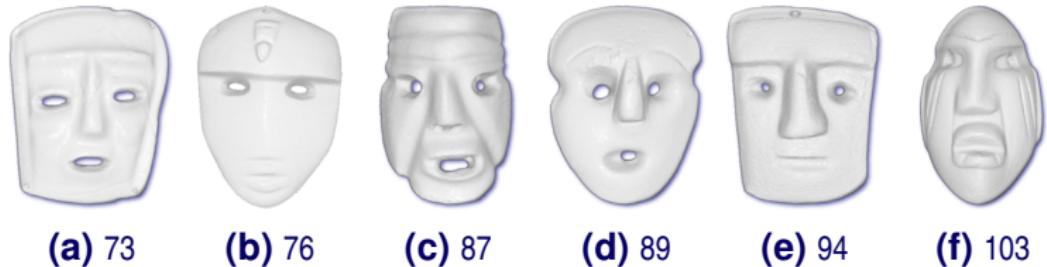
# Fam 02 — Cilindros $T = 32$



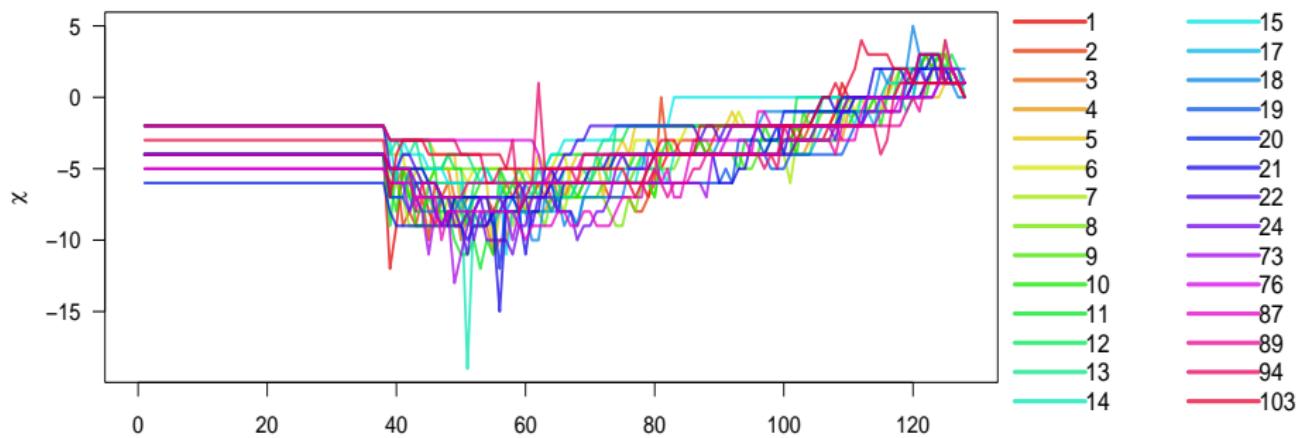
ECG – cyl32 – fam 2



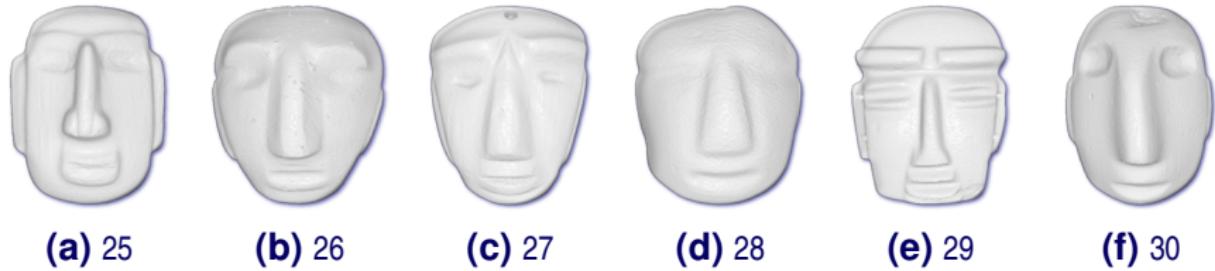
# Fam 02 — Esfera $T = 128$



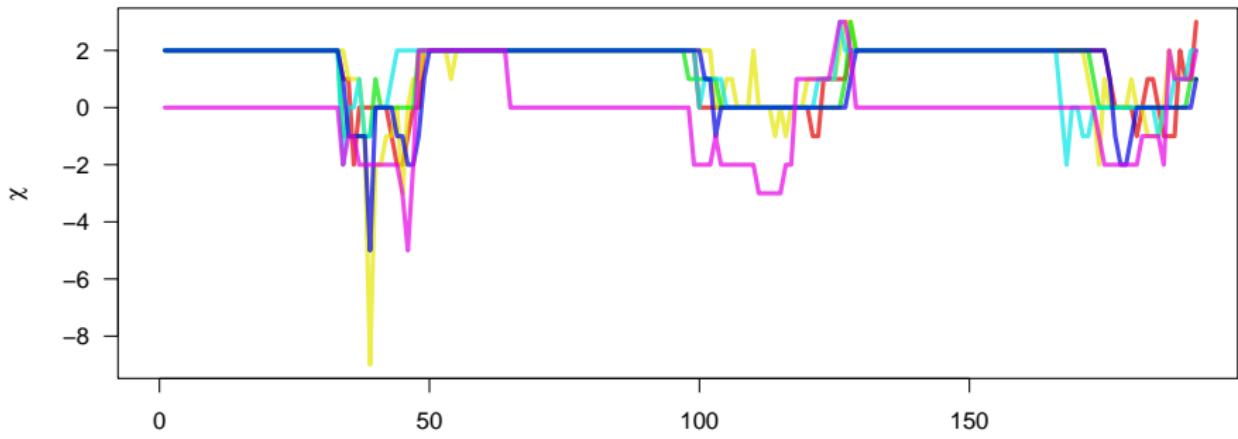
ECG - sphsqrt128 - fam 2



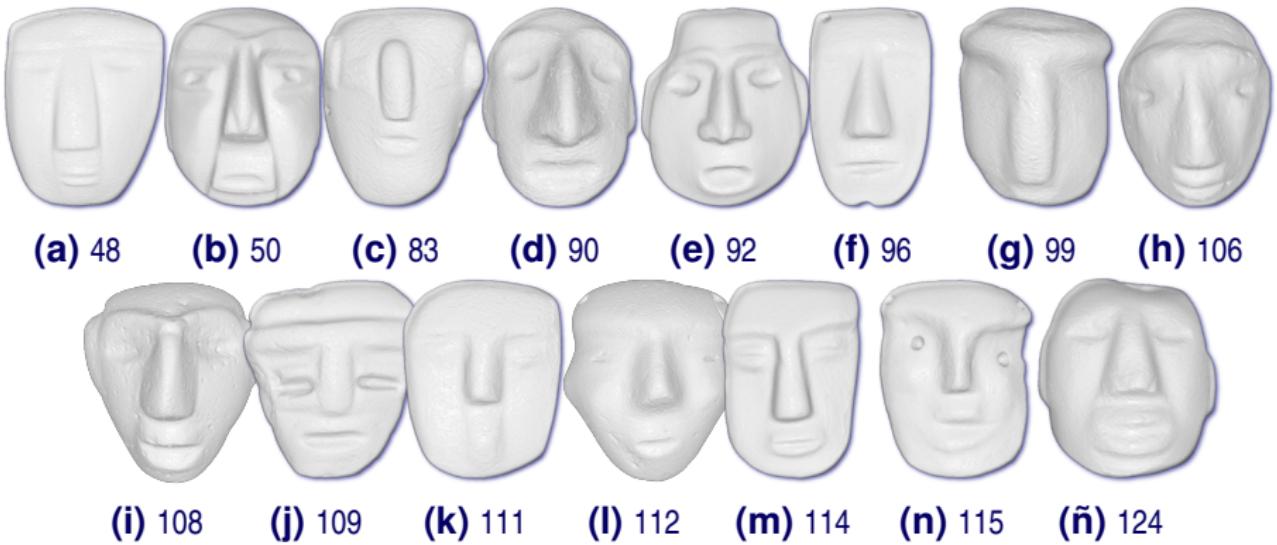
# Familia 03



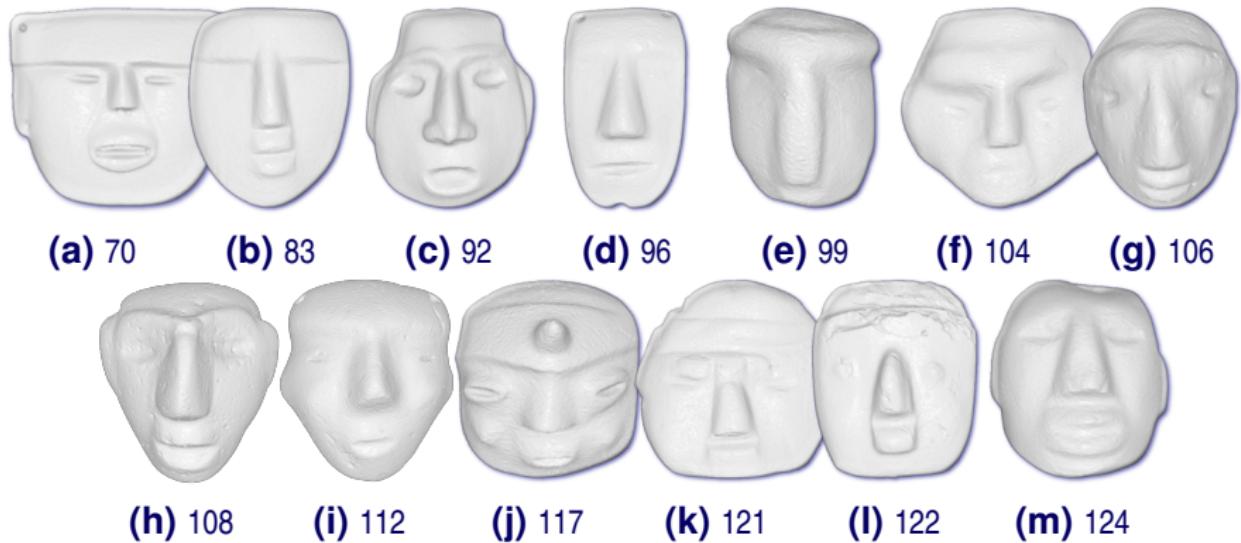
ECG = cyl2 – T = 64 – set = 3



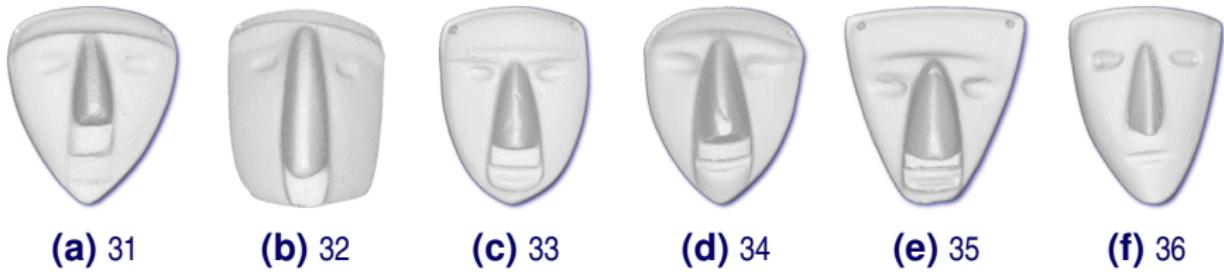
# Cilindros T=32



# Esferas $T = 128$

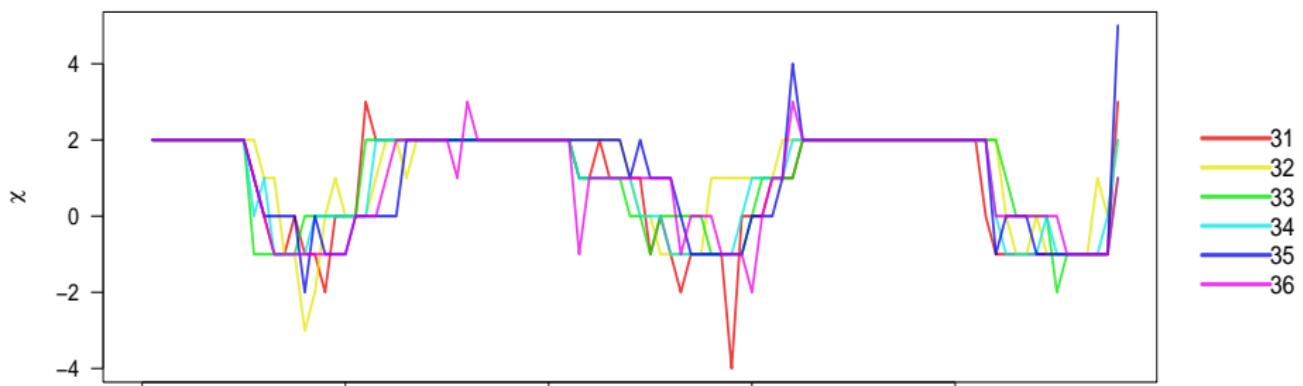


# Familia 04



Masks in the original set 04

ECG - cyl32 - fam 4



# Familia 05



(a) 37



(b) 38



(c) 39



(d) 40



(e) 41



(f) 42



(g) 43



(h) 44



(i) 45



(j) 46



(k) 47



(l) 48



(m) 49



(n) 50



(ñ) 51



(o) 52



(p) 53

# Cilindros $T = 32$



(a) 52

(b) 78

(c) 81

(d) 82

(e) 93

(f) 97



(g) 101



(h) 105



(i) 113



(j) 120



(k) 128

# Familia 07



(a) 56



(b) 57

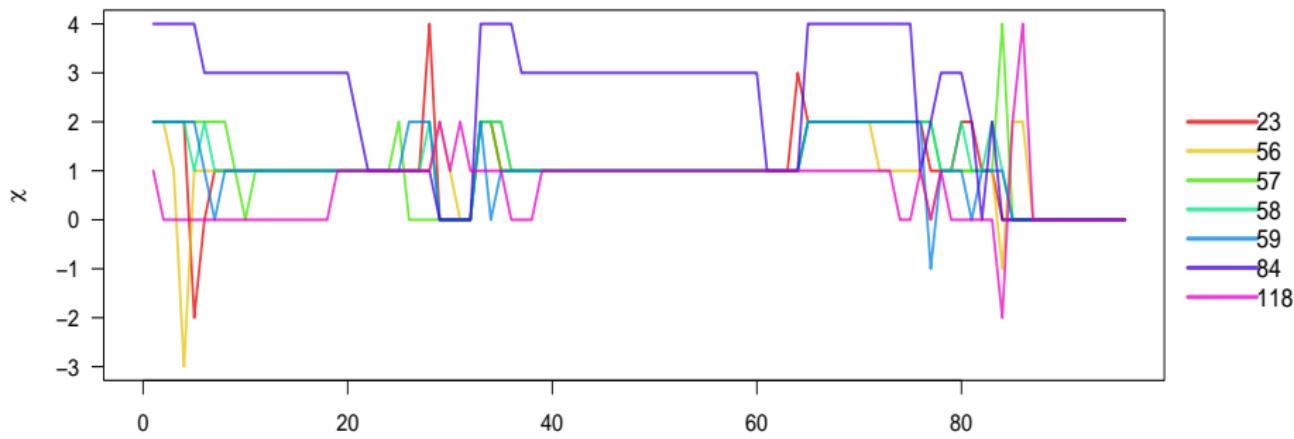


(c) 58

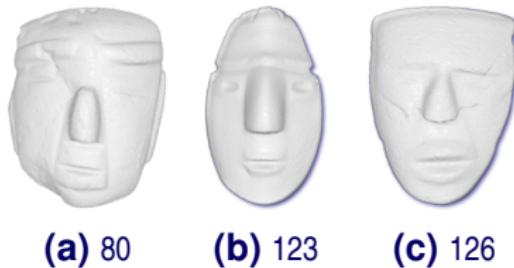


(d) 59

ECG – planar32 – fam 7

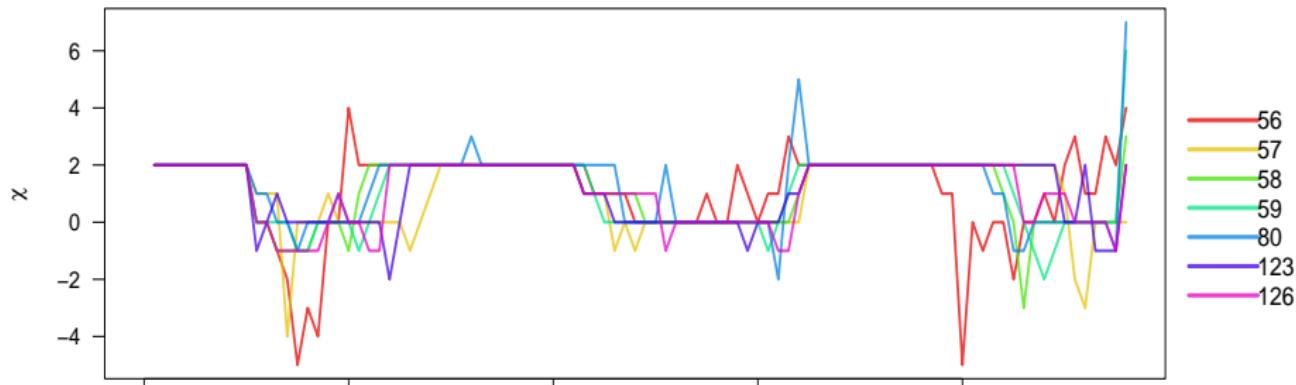


# Cilindros $T = 32$



Masks assigned to Set 07 after running 72 polynomial SVMs

ECG - cyl32 - fam 7



# Esferas $T = 128$



(a) 48

(b) 50

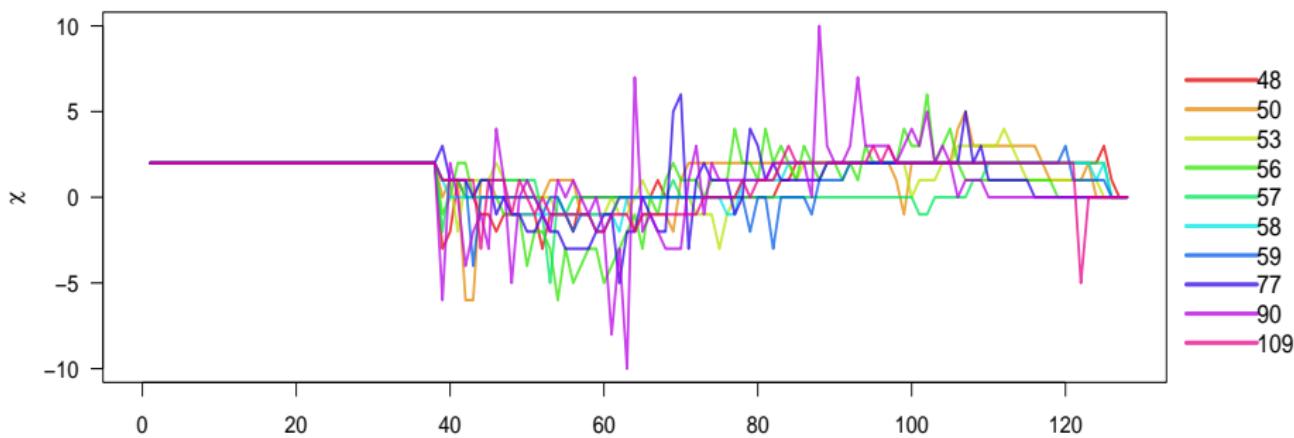
(c) 53

(d) 77

(e) 90

(f) 109

ECG - sphsqrt128 - fam 7



# Comentarios Finales

- El cómputo de la GCE es una operación sencilla de complejidad y memoria lineales. Procesa rápidamente objetos de decenas de miles de vértices.
- Se puede explorar con muchas más variables.
- Una mayor cantidad de máscaras puede proveer de mejores conjuntos de entrenamientos y por ende, mejores clasificaciones.
- Más especímenes permitirán también experimentar con métodos de clasificación no supervisada.

# Referencias

-  M. A. Armstrong *Basic Topology* Springer-Verlag (1983), New York.
-  C. Burges "A Tutorial on Support Vector Machines for Pattern Recognition". *Data Mining and Knowledge Discovery* Vol.2 (1998) pp.121-167.  
<https://doi.org/10.1023/A:1009715923555>
-  D. Jiménez Badillo, S. Ruíz Correa, O. Mendoza Montoya *Analyzing formal features of archaeological artefacts through the application of Spectral Clustering*. Conferencia del Digital Classicist Seminar (06/11/2012). Deutsches Archäologisches Institut, Berlín. <http://hdl.handle.net/11858/00-1780-0000-000B-216A-E>
-  E. Richardson, M. Weirman, "Efficient classification using the Euler Characteristic". *Pattern Recognition Letters* Vol.49, (2014) pp.99-106.  
<http://www.sciencedirect.com/science/article/pii/S0167865514002050>

- TAREA: Calcular GCEs de la siguiente máscara



Thank God  
It's Friday!

# Agradecimientos



CIMAT



IPICYT



CONACYT

Consejo Nacional de Ciencia y Tecnología



Instituto Nacional  
de Antropología  
e Historia