

Reconocimiento de Gestos Dinámicos

Facundo Quiroga

Directora: Lic. Laura Lanzarini

Co-director: Lic. Leonardo Corbalán

5 de marzo de 2014



¿Reconocimiento de gestos?



La mayor parte de la comunicación es no verbal



Lesiones por sobreuso de PCs



Método de interacción

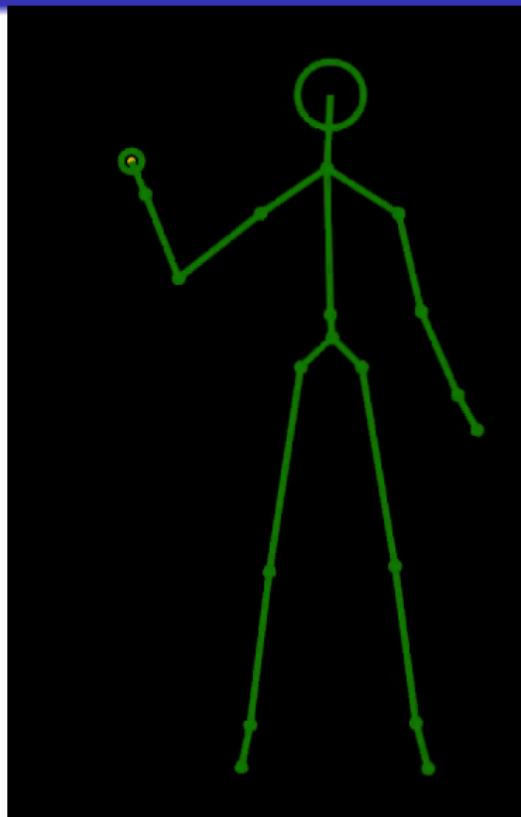


Existen varios tipos de gestos y formas de reconocimiento.

Problema

Descripción del problema

- Reconocimiento de gestos con la mano en 3D
 - Definidos por el usuario
 - Pocos gestos para entrenar (3 por cada **clase de gesto**)



Modelos explícitos

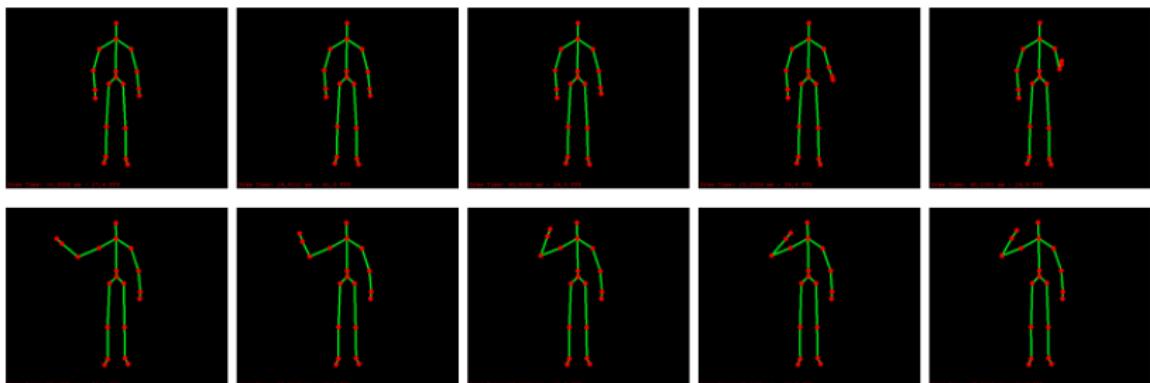
Gesto de saludo

- La mano comienza debajo del pecho
 - Subir la mano en línea recta a una velocidad mayor a $V \frac{m}{seg}$
 - Detener la mano abruptamente arriba de la cabeza

En GestureML:

```
<inertial_filter>
  <property ref="drag_dy" active="true" friction="0.9"/>
</inertial_filter>
<delta_filter>
  <property ref="drag_dy" active="true" delta_min="0.5"
            delta_max="500"/>
</delta_filter>
```

Modelos generados por Aprendizaje Automático



Entrenamiento

Modelo

Modelos Explícitos vs Aprendizaje Automático

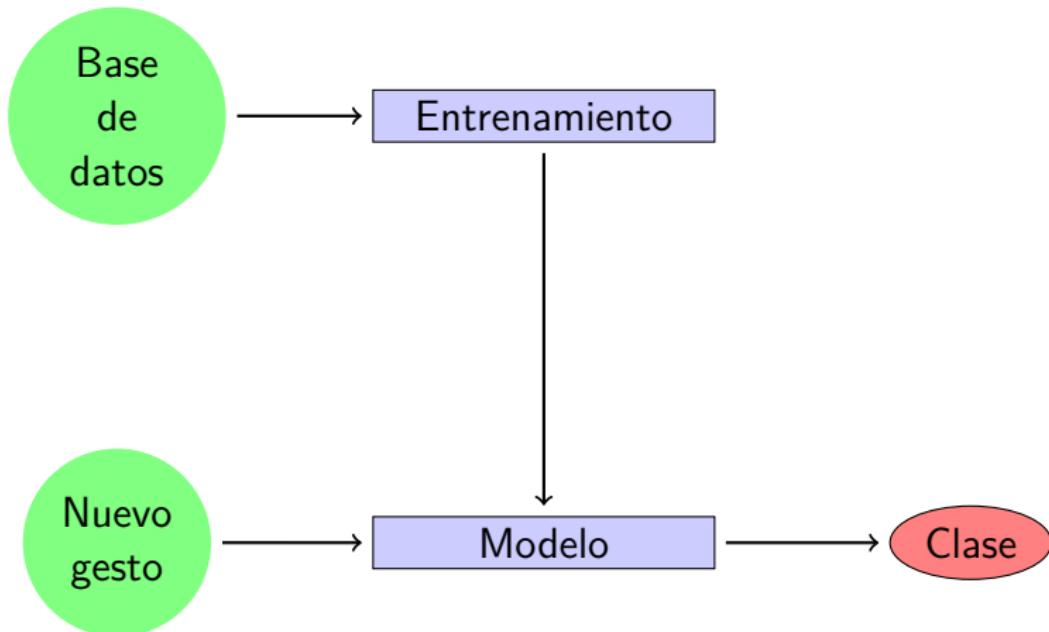
Definir un modelo explícitamente

- *Lenguaje de definición de gestos*
- Necesita conocimiento experto
- Difícil agregar nuevos gestos
- Vocabulario limitado
- Poco adaptables a distintas personas

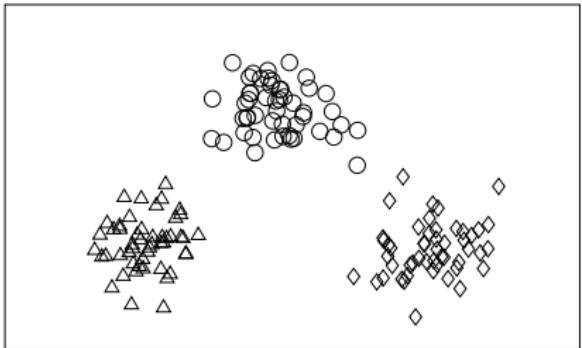
Entrenar un modelo con Aprendizaje Automático

- *Grabar una base datos de ejemplos para cada tipo de gesto y entrenar un modelo*
- Los usuarios pueden grabar nuevos gestos fácilmente
- Puede mejorarse automáticamente al usar el sistema
- No hay algoritmos de 99.9 % de efectividad

Aprendizaje automático

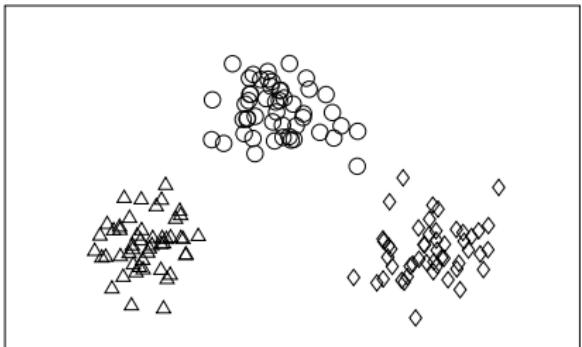


Entrenamiento + Clasificación

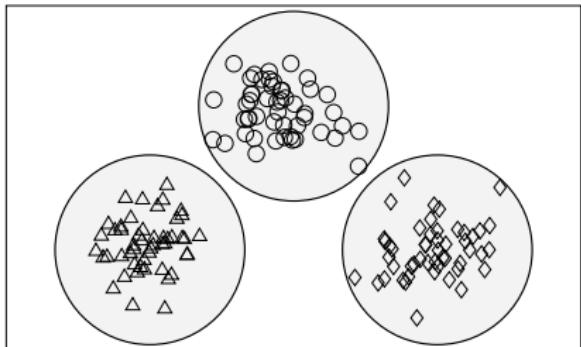


BD

Entrenamiento + Clasificación

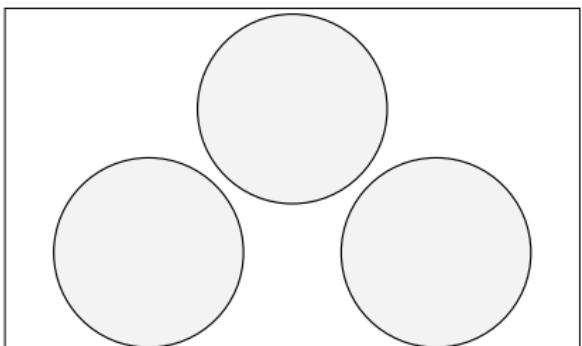
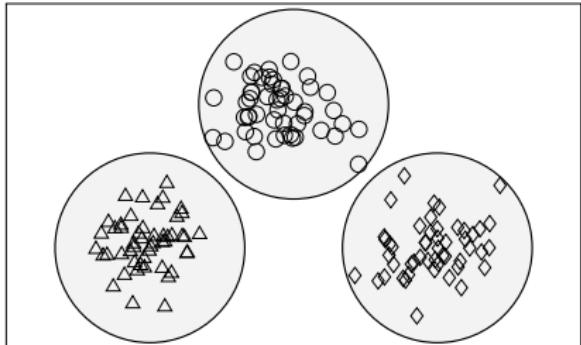
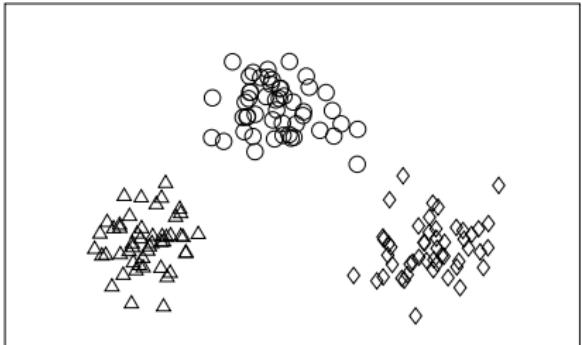


BD

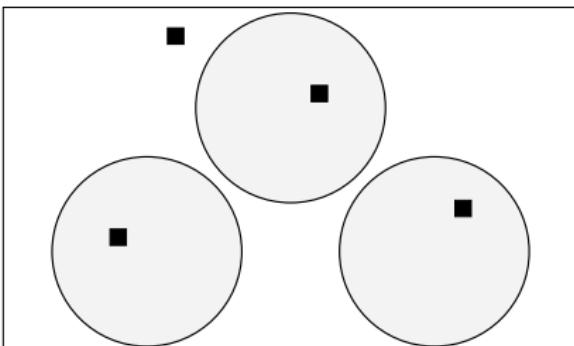
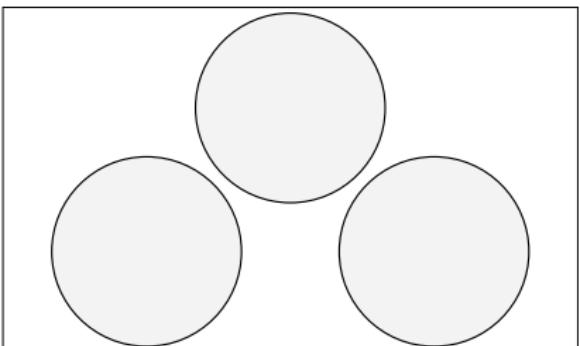
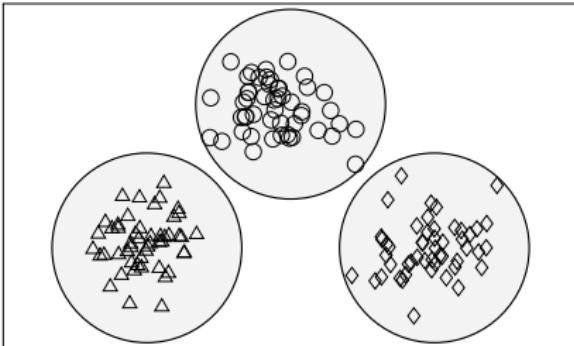
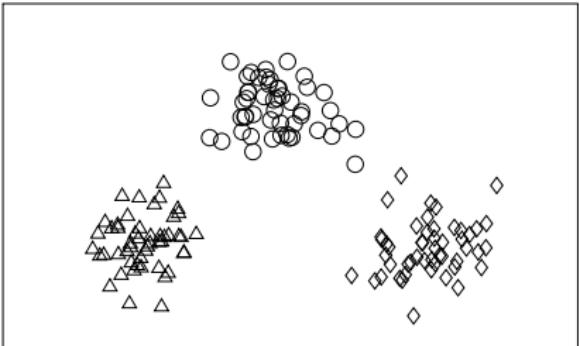


Entrenamiento

Entrenamiento + Clasificación



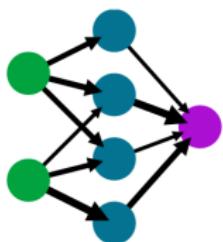
Entrenamiento + Clasificación



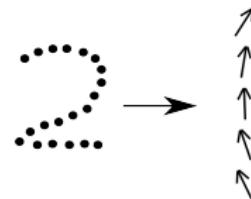
Trabajo realizado



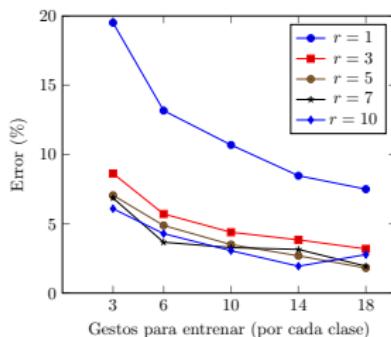
Base de datos de gestos



Clasificador Neuronal
Competitivo (CNC)



Modelo y representaciones de
gestos

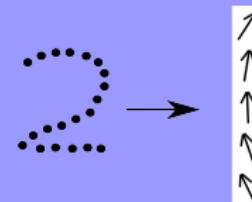


Resultados

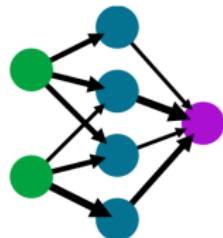
Gestos y Bases de datos



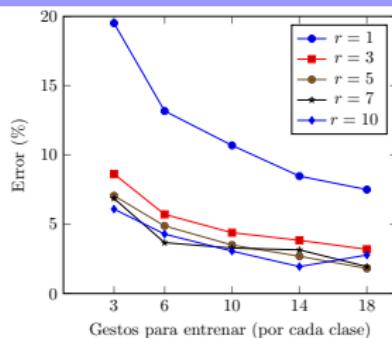
Base de datos de gestos



Modelo y representaciones de gestos



Clasificador Neuronal Competitivo (CNC)



Resultados

Dispositivo Kinect



Sensor de profundidad
(emisor +
receptor IR)

Imagen RGB normal



Kinect SDK



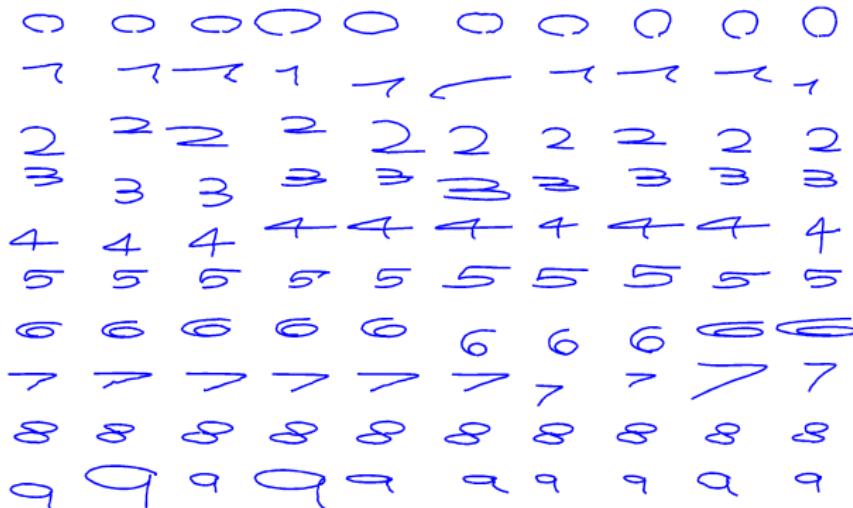
=



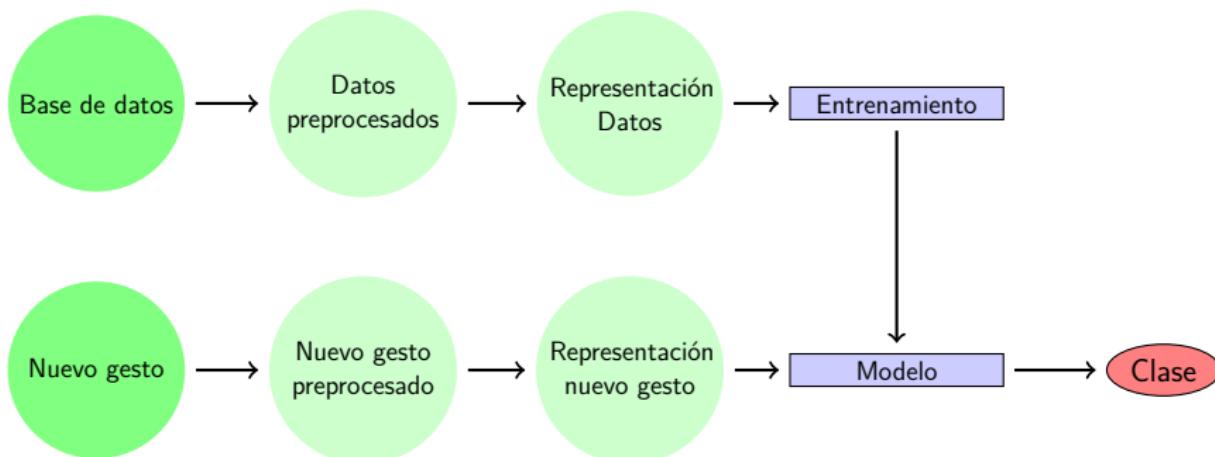
- Hasta 5 personas
- Tiempo real, 30 fps
- Coordenadas espacio 3D
- Posición 20 articulaciones

Base de datos LNHG

720 gestos 36 clases 20 ejemplares por clase

(Proyecciones en el plano $X - Y$)

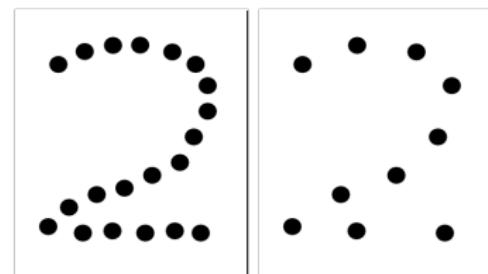
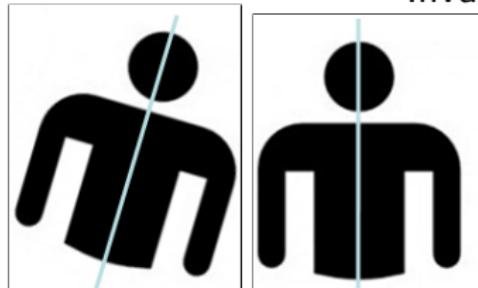
Preprocesamiento y Representación



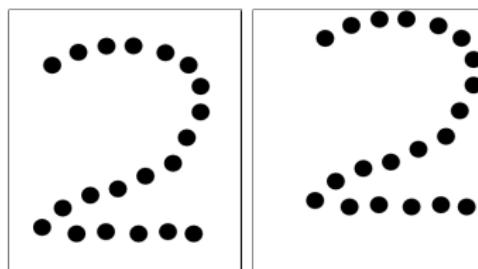
Modelo de gestos

Gesto = Trayectoria en el espacio virtual

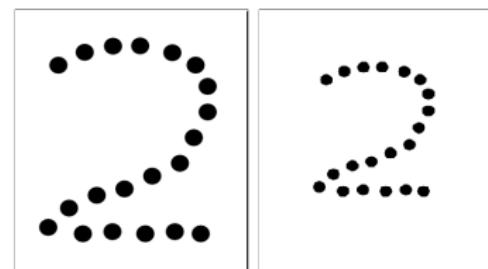
Invariante a:



Rotación del usuario



Velocidad

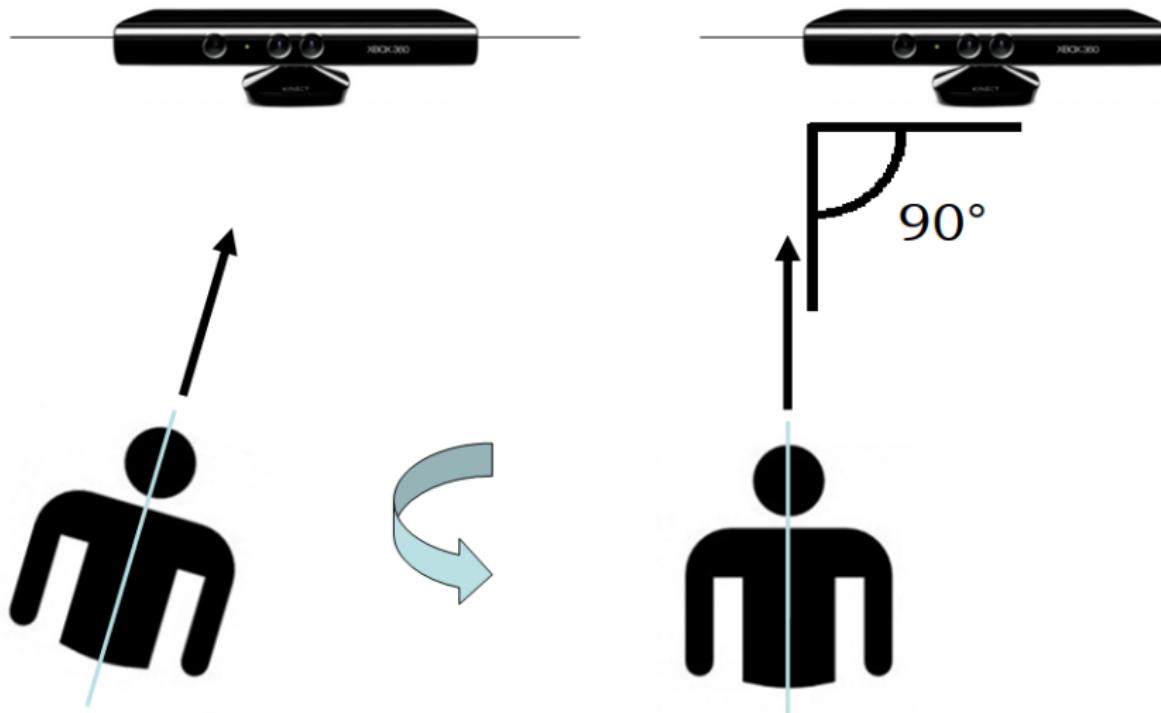


Traslación

Escala

Preprocesamiento y Representación

Preprocesamiento: Rotación

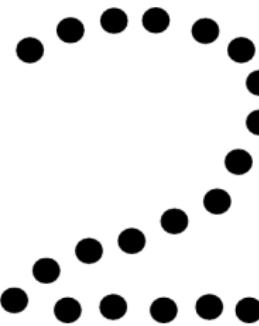


Pre: Suavizado y re-muestreo → Invariancia a velocidad

Ejemplo de re-muestreo con $n = 6$

- Longitud de arco $L = 10$.
- Nuevos puntos de muestreo: $d_k = Lk/(n - 1) = 10k/5 = 2k$
 $k = 0, \dots, (n - 1)$. ($d = 0, 2, 4, 6, 8, 10$)
- Calcular el punto a distancia d_k del principio, interpolando.

Representacion: Secuencia de Direcciones



Primera diferencia ($d = g[n + 1] - g[n]$) → Invariancia a la traslación



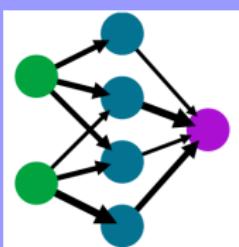
Normalización ($\|d\| = 1$) → Invariancia a la escala



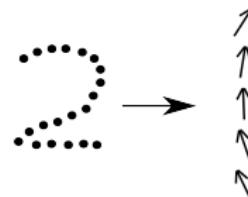
Clasificador Neuronal Competitivo



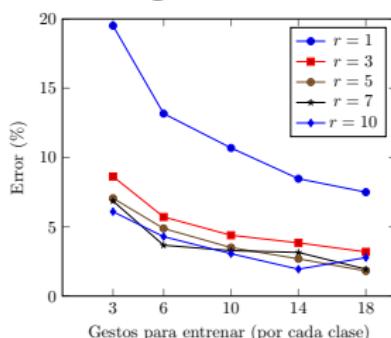
Base de datos de gestos



Clasificador Neuronal
Competitivo (CNC)

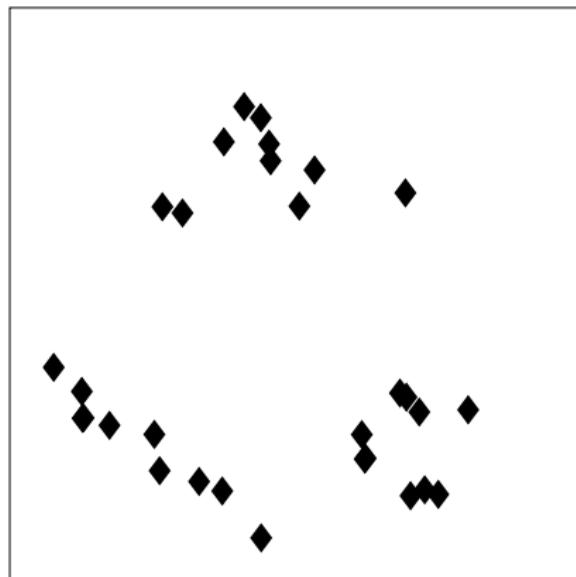
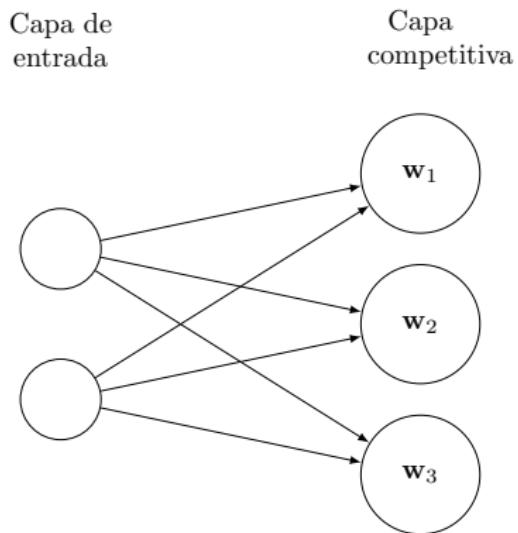


Modelo y representaciones de
gestos



Resultados

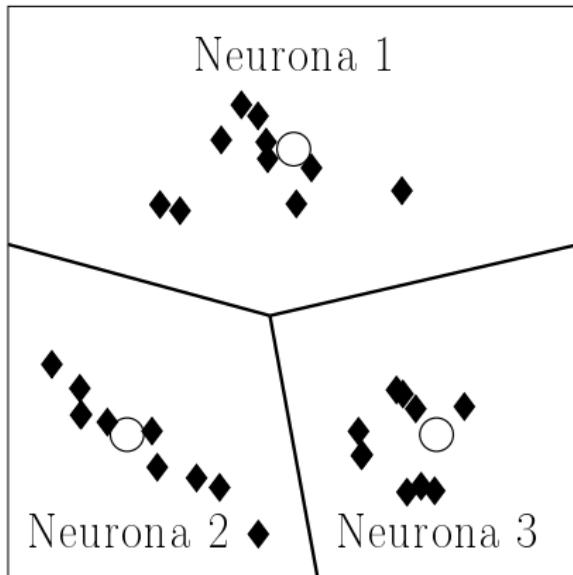
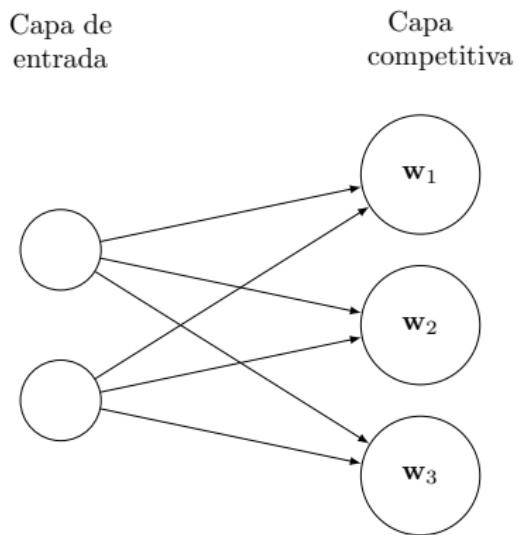
Redes neuronales competitivas



Características

- Clustering, no clasificación
- Modelo base del CNC

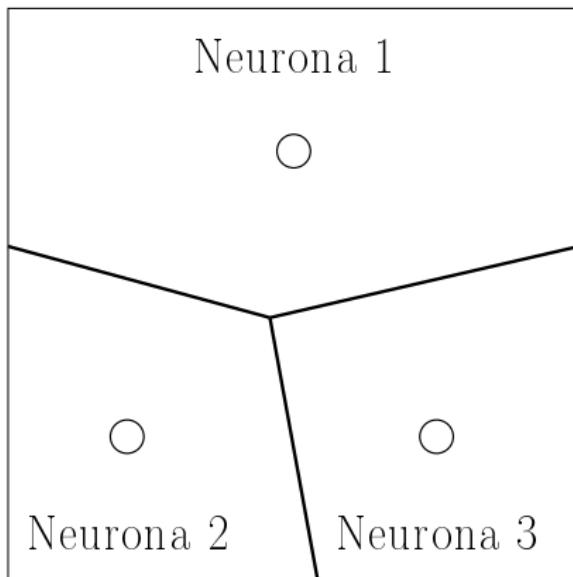
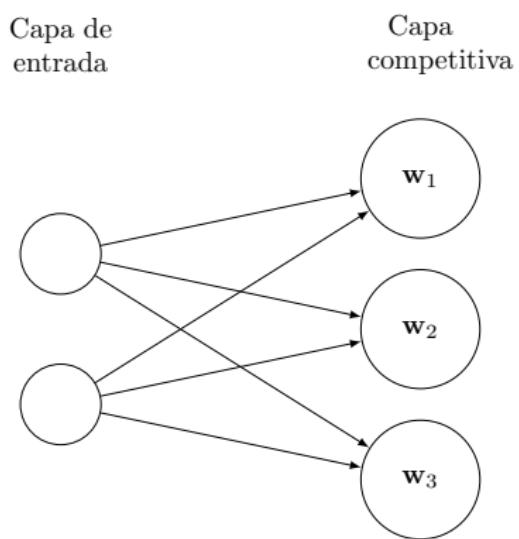
Redes neuronales competitivas



Características

- Clustering, no clasificación
 - Modelo base del CNC

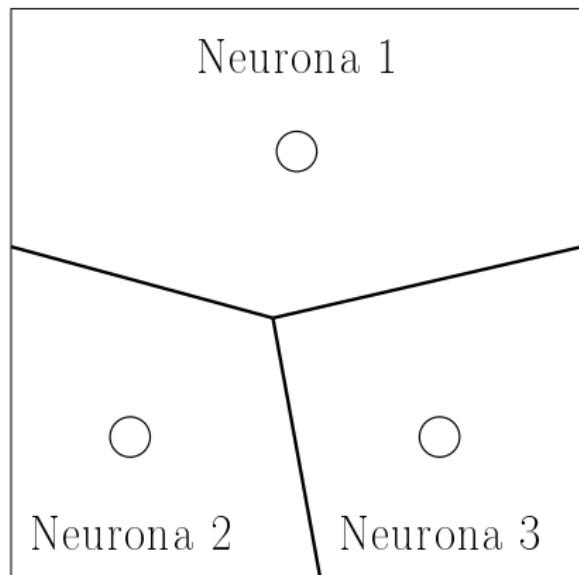
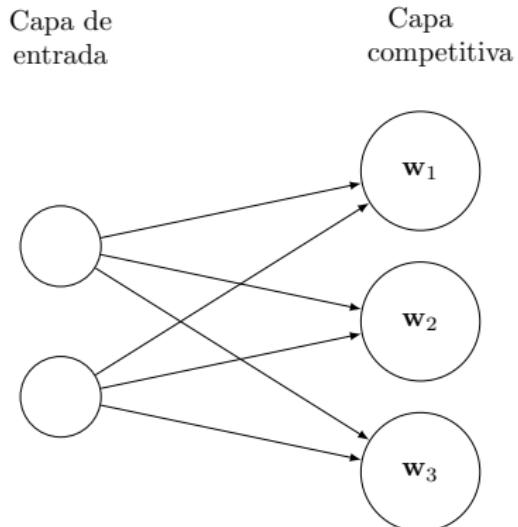
Redes neuronales competitivas



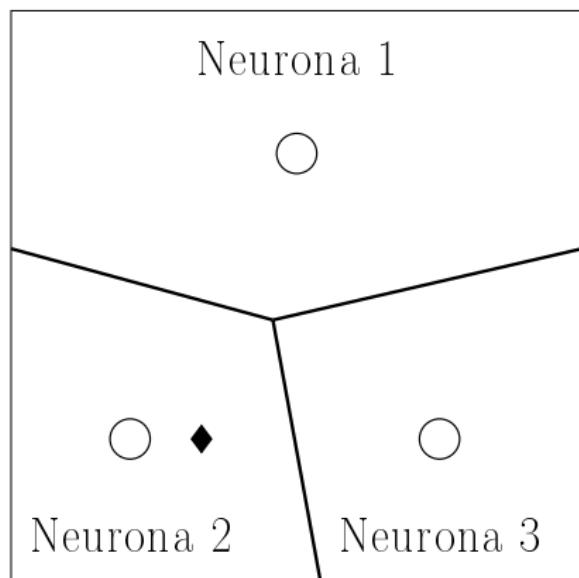
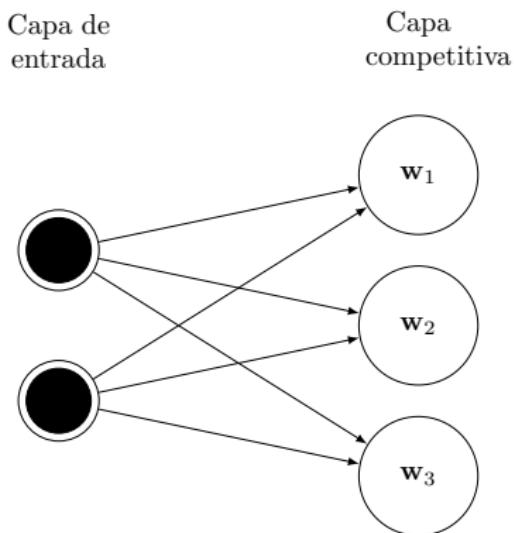
Características

- Clustering, no clasificación
 - Modelo base del CNC

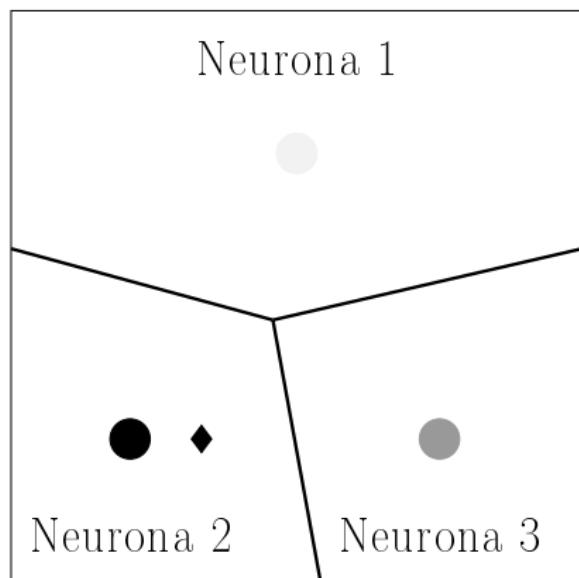
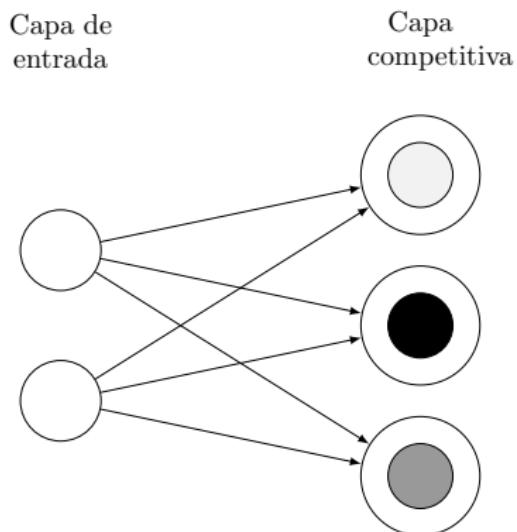
Funcionamiento: Clustering



Funcionamiento: Clustering



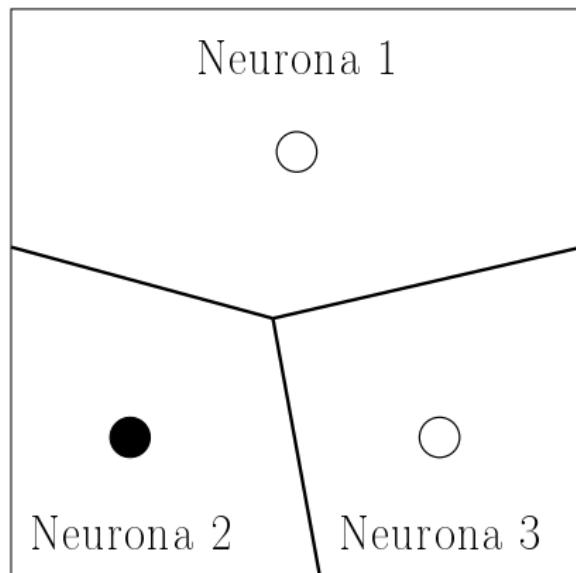
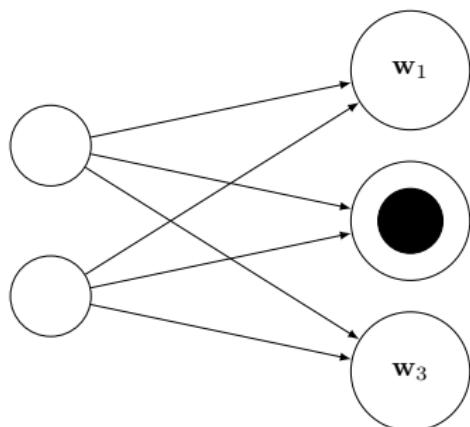
Funcionamiento: Clustering



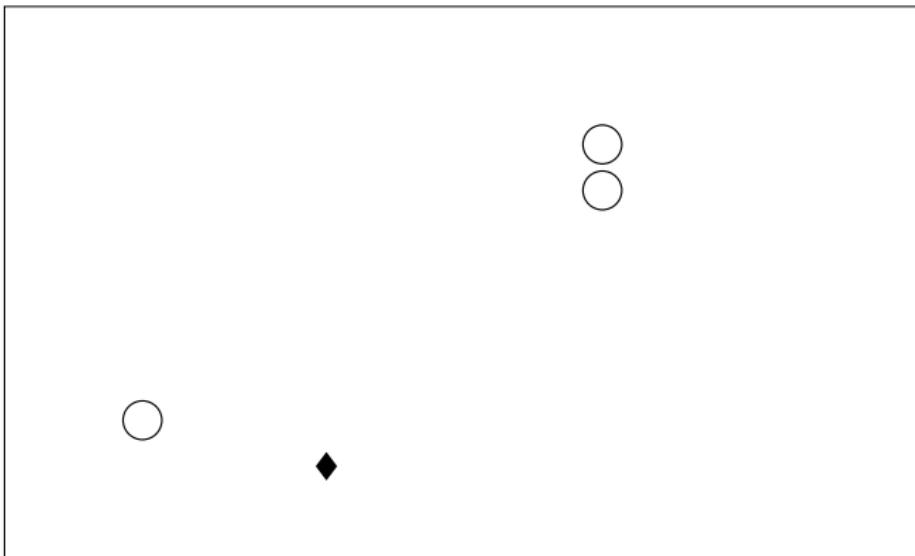
Funcionamiento: Clustering

Capa de entrada

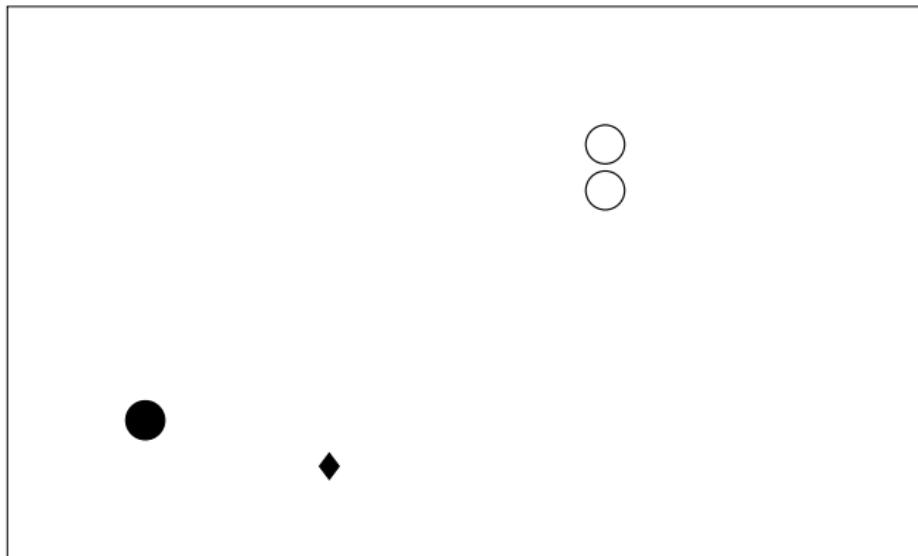
Capa competitiva



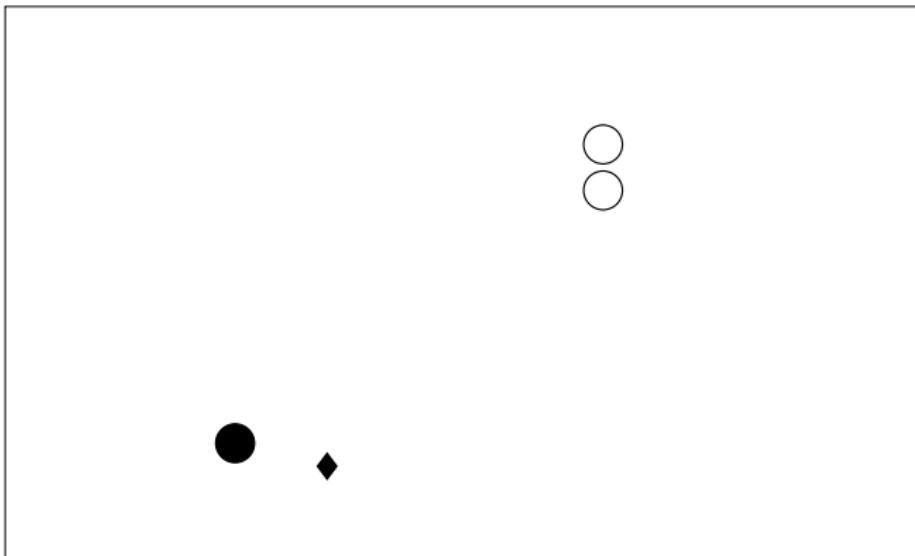
Entrenamiento (de a uno)



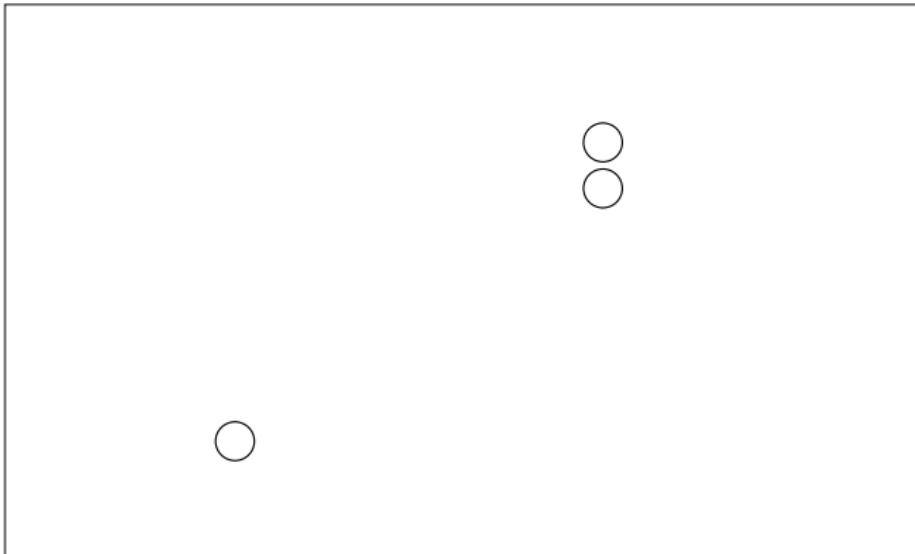
Entrenamiento (de a uno)



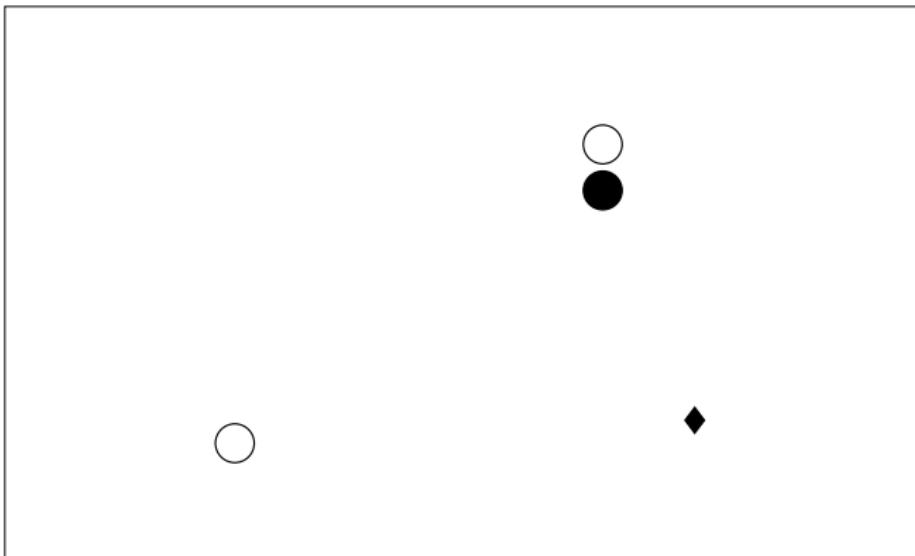
Entrenamiento (de a uno)



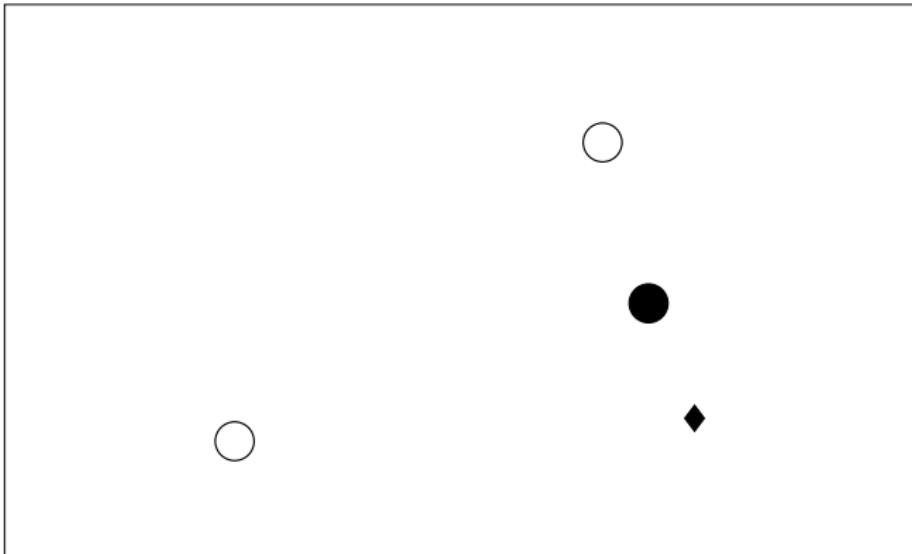
Entrenamiento (de a uno)



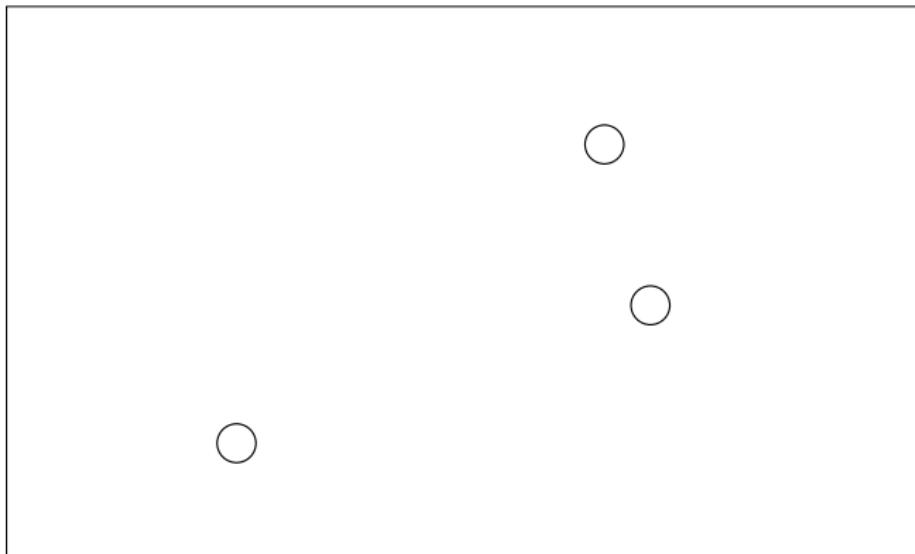
Entrenamiento (de a uno)



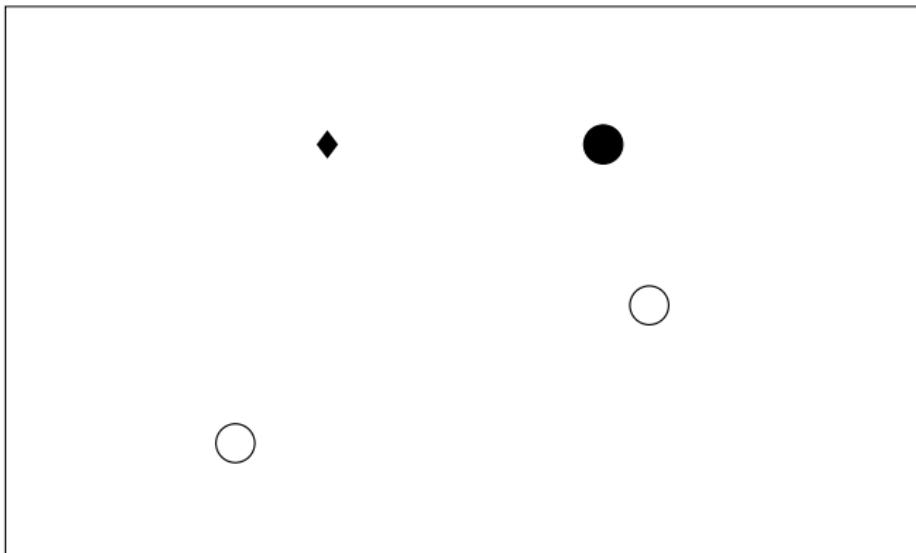
Entrenamiento (de a uno)



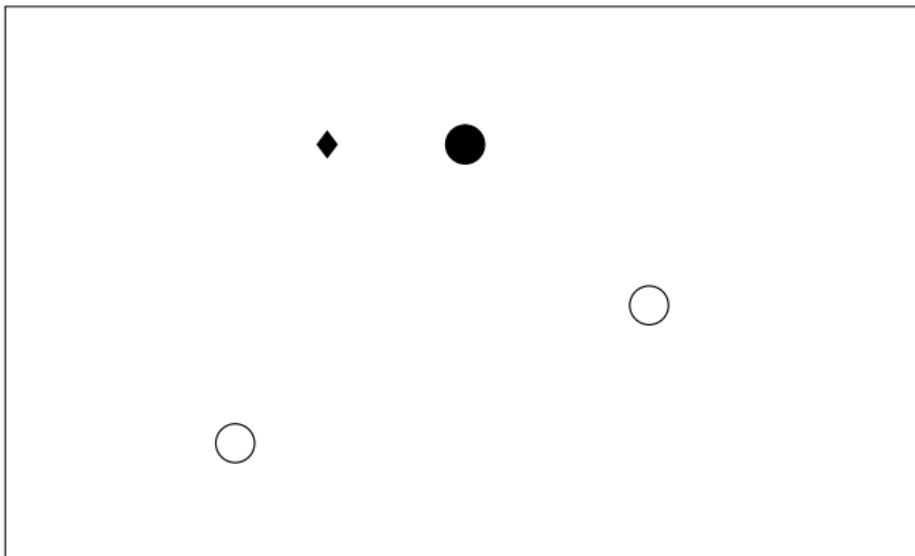
Entrenamiento (de a uno)



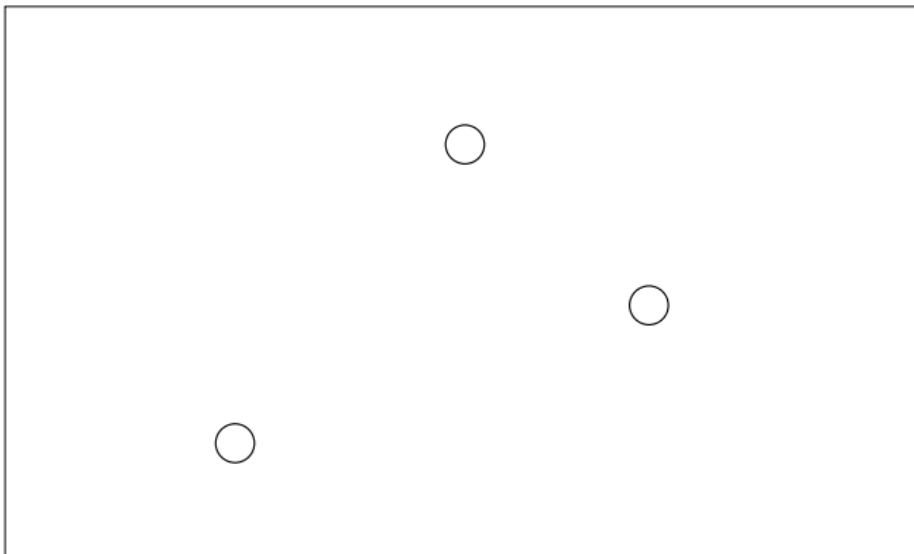
Entrenamiento (de a uno)



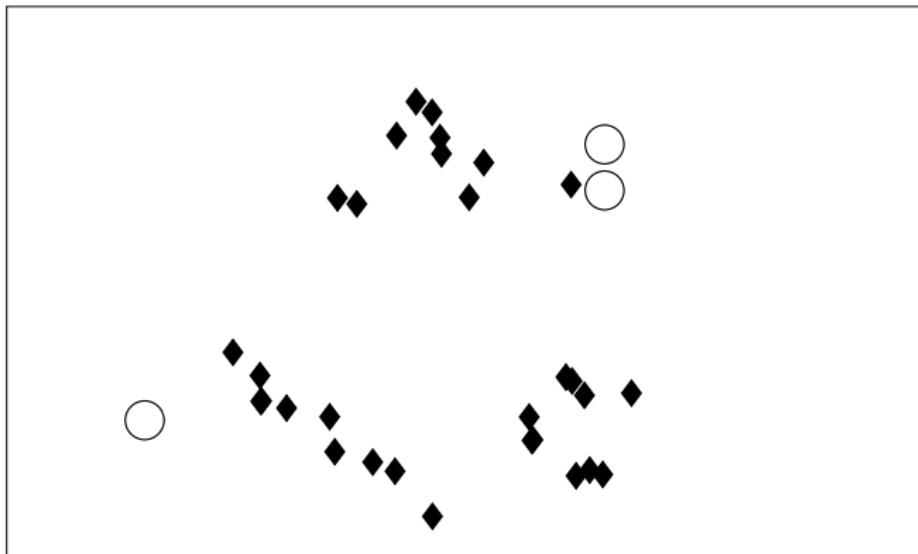
Entrenamiento (de a uno)



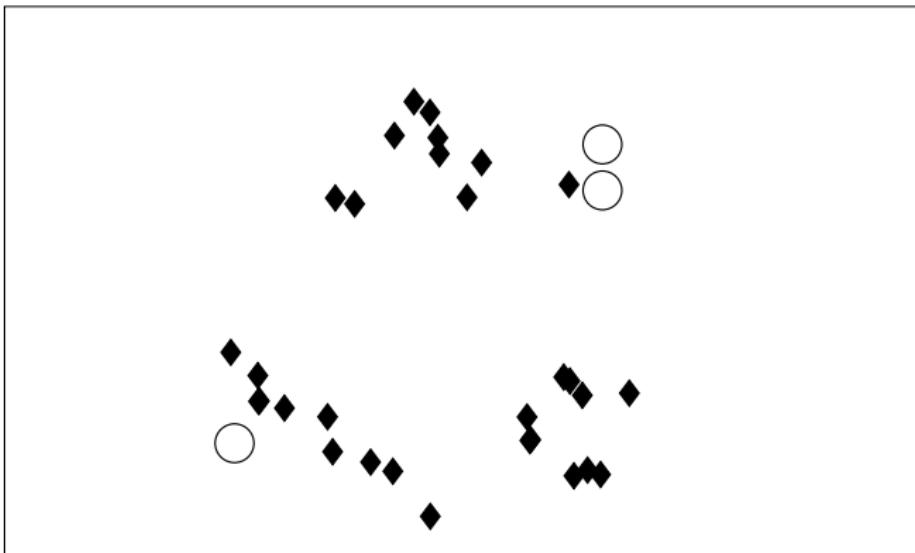
Entrenamiento (de a uno)



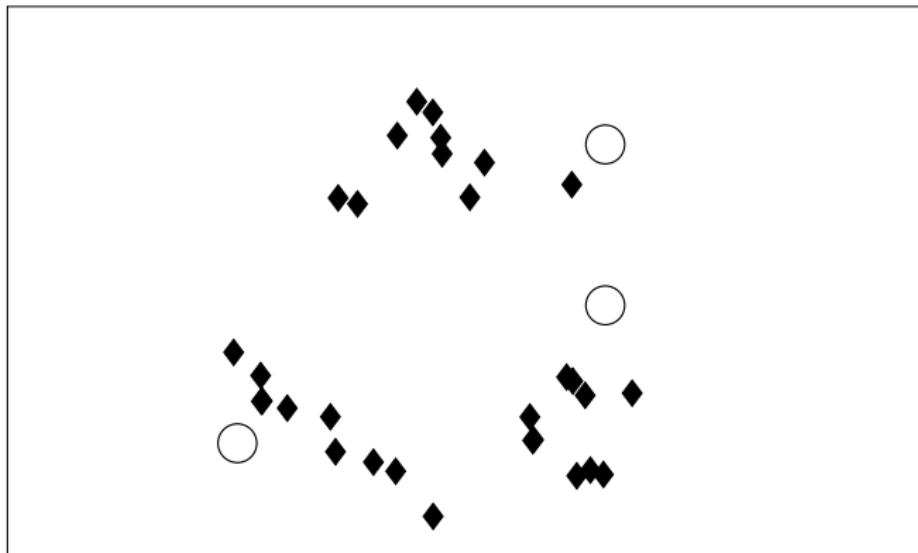
Entrenamiento (Todos los ejemplares)



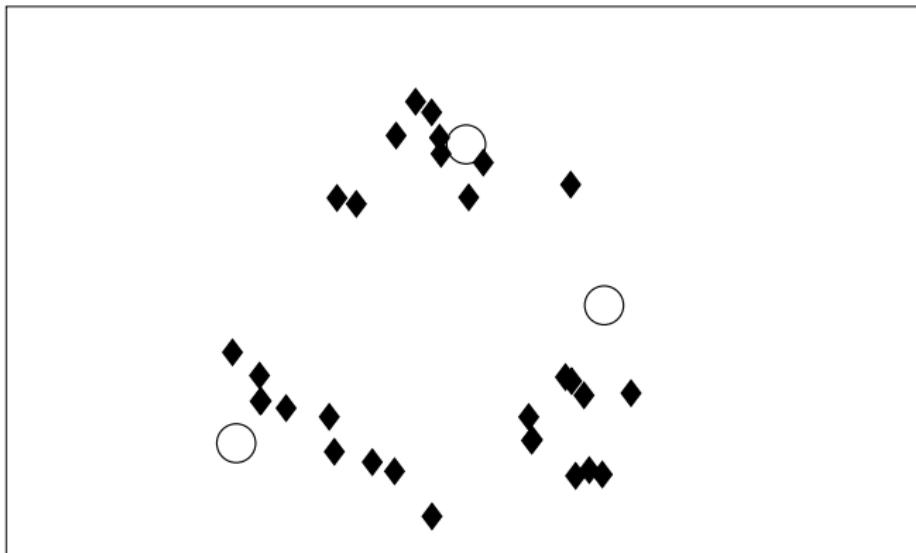
Entrenamiento (Todos los ejemplares)



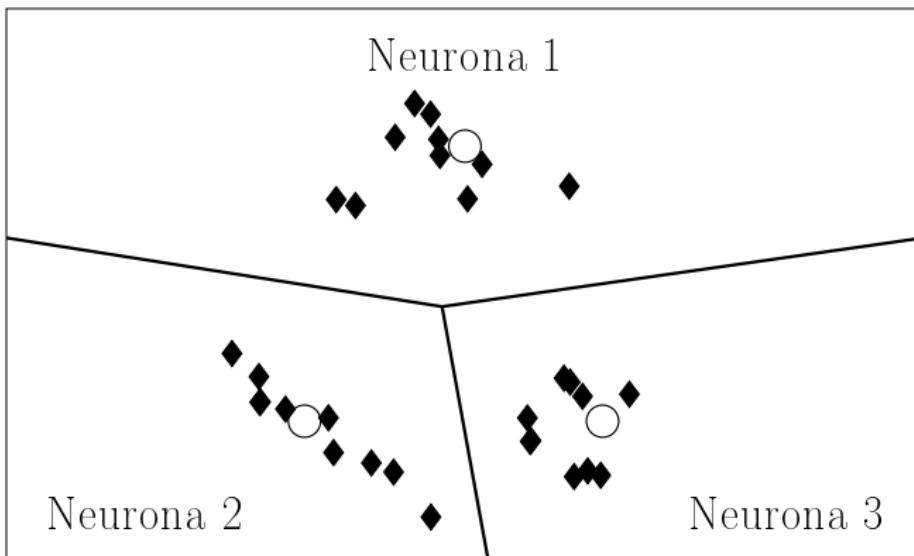
Entrenamiento (Todos los ejemplares)



Entrenamiento (Todos los ejemplares)

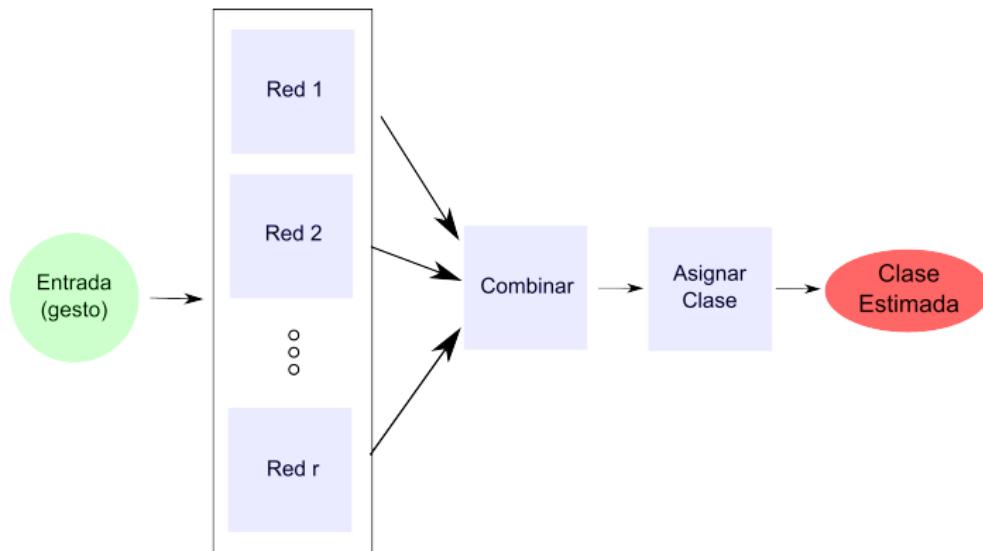


Entrenamiento (Todos los ejemplares)



Clasificador neuronal competitivo (CNC)

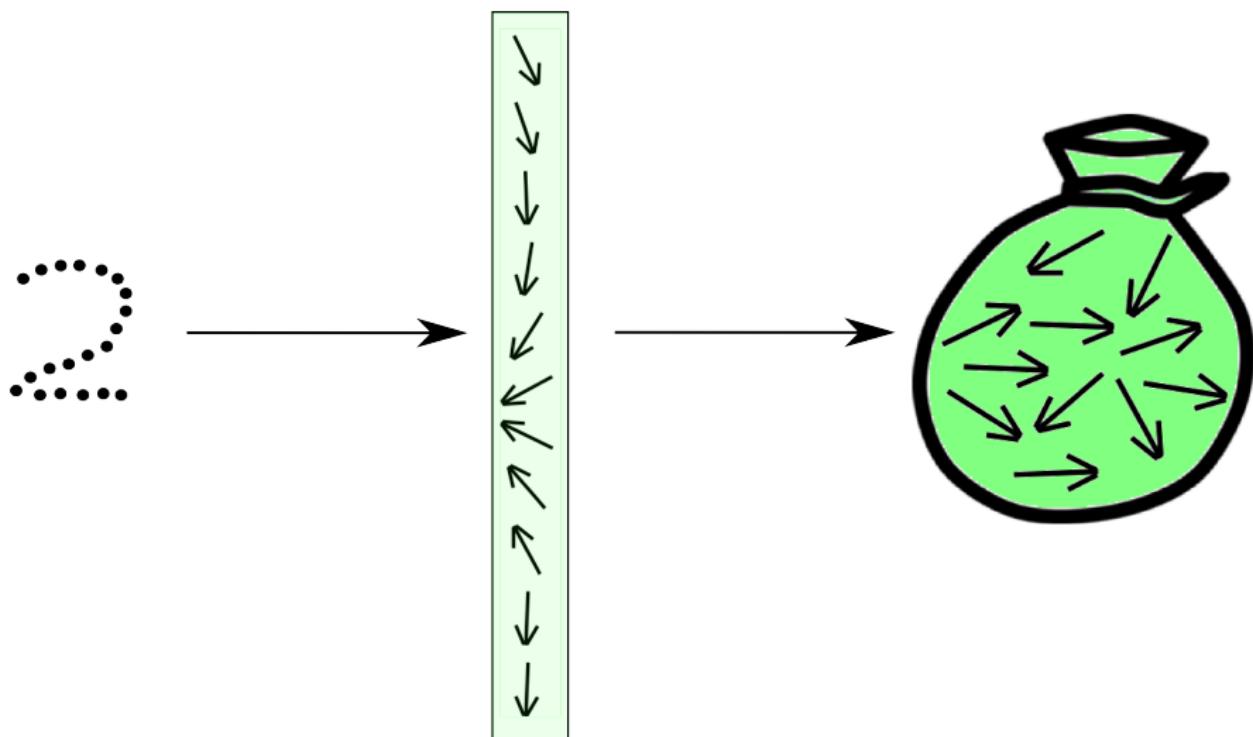
Clasificador neuronal competitivo (CNC)



- Basado en redes neuronales competitivas
- Buen rendimiento aún con pocos gestos para entrenar
- Esquema de **bagging**
- Puede usarse sin re-muestrear los gestos.

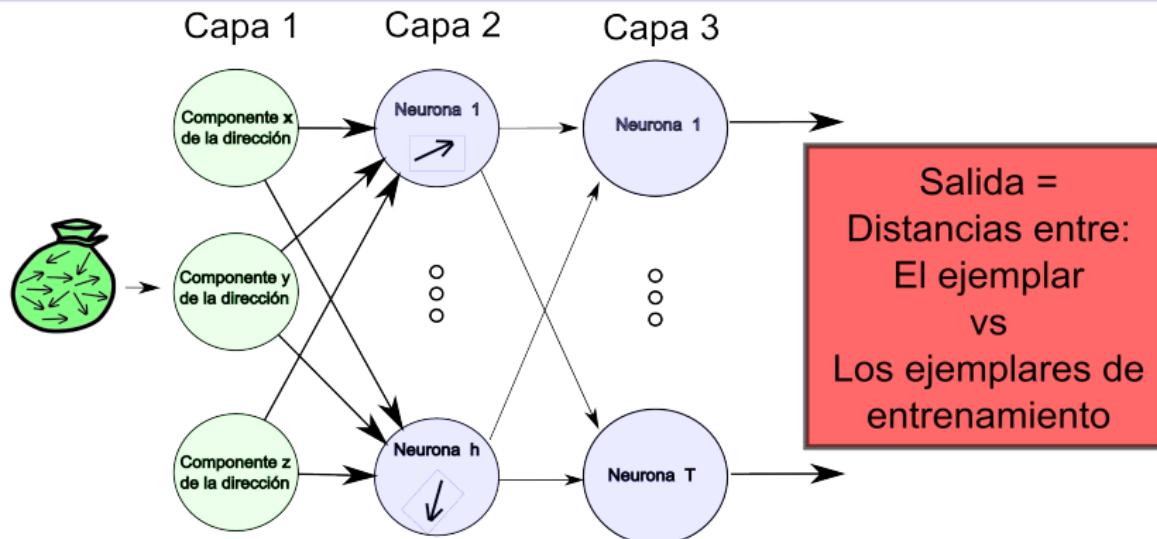
CNC: Funcionamiento de cada red

Idea: Descartar secuencia de direcciones



CNC: Funcionamiento de cada red

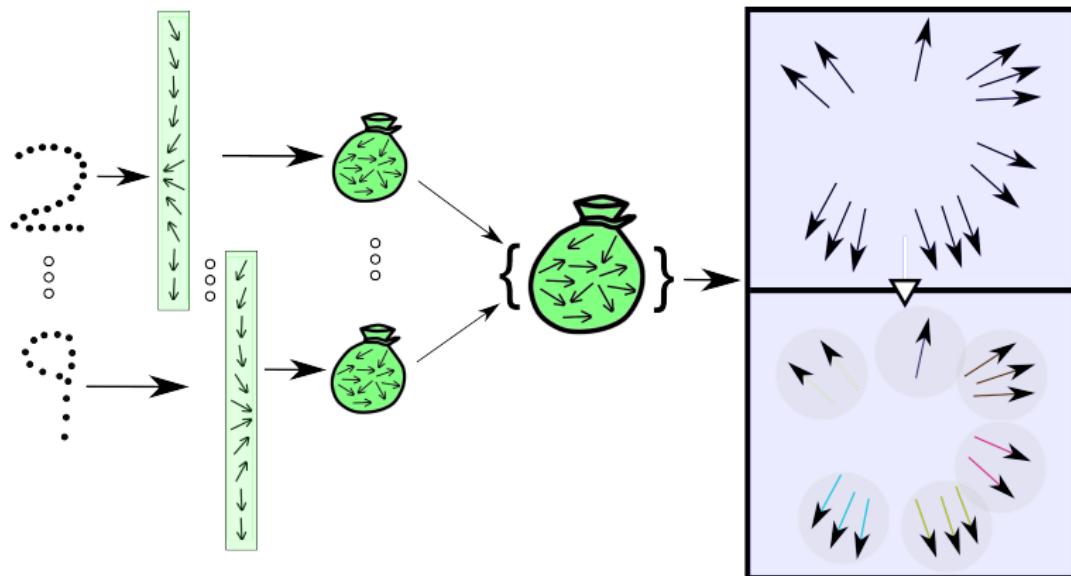
Estructura de cada red



- h = Cantidad de neuronas de la capa 2, \rightarrow parámetro a seleccionar
- T = Cantidad de gestos del conjunto de entrenamiento

CNC: Funcionamiento de cada red

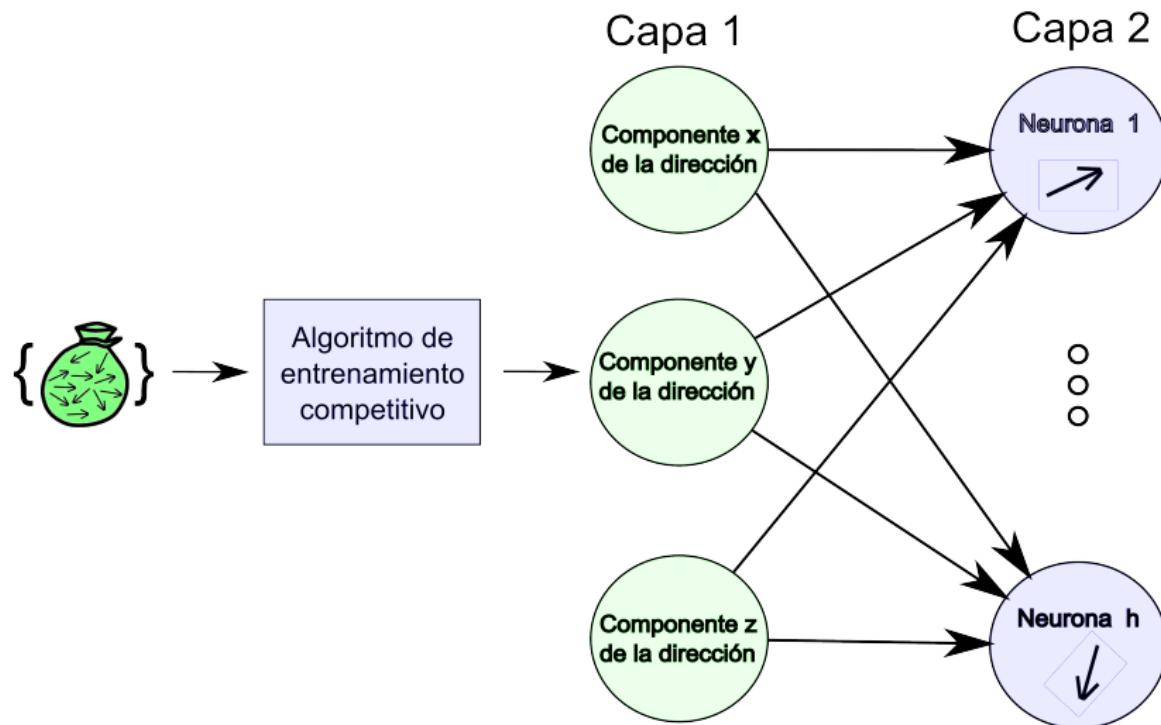
Idea: Clustering de direcciones



- $h =$ Cantidad de neuronas de la capa 2, \rightarrow parámetro a seleccionar

CNC: Funcionamiento de cada red

Entrenamiento capa 2



CNC: Funcionamiento de cada red

Idea: Codificación de gestos: Ejemplo



Gesto x: → → → → ↘ ↘ ↙ ↙ → →

Clustering: ↑ → ↓ ↓ ↑ ↓
1 2 3 4 5 6

Gesto v:

CNC: Funcionamiento de cada red

Idea: Codificación de gestos: Ejemplo



Gesto x: ↗ ↗ ↗ ↗ ↘ ↘ ↙ ↙ ↙ ↗ ↗
1 1 2 2 3 4 6 6 6 2 2

Clustering: ↗ → ↘ ↓ ↙ ↙
1 2 3 4 5 6

Gesto v:

CNC: Funcionamiento de cada red

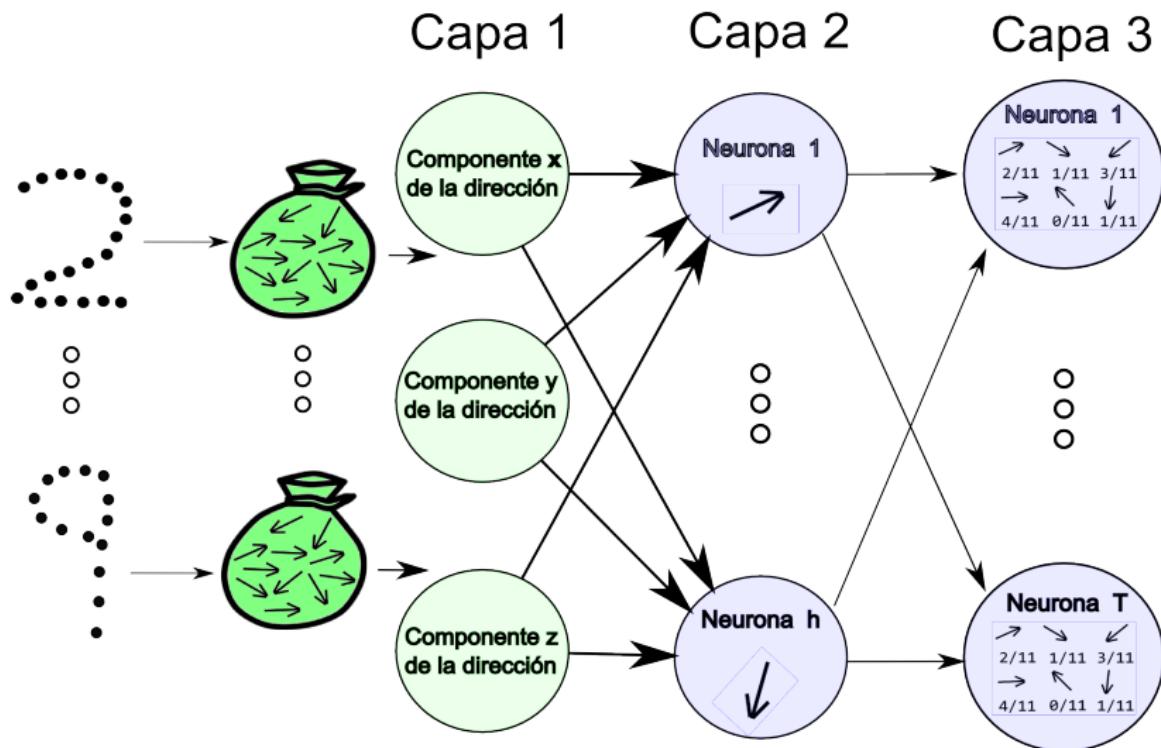
Idea: Codificación de gestos: Ejemplo



Gesto x:	↗	↗	↗	↗	↘	↙	↙	↖	↖	↗	↗
	1	1	2	2	3	4	6	6	6	2	2
Clustering:	↗	→	↘	↓	↙	↖	↖	↗	↗	↘	↘
	1	2	3	4	5	6					
Gesto v:	↗	→	↘	↓	↖	↖					
	2/11	4/11	1/11	1/11	0/11	3/11					

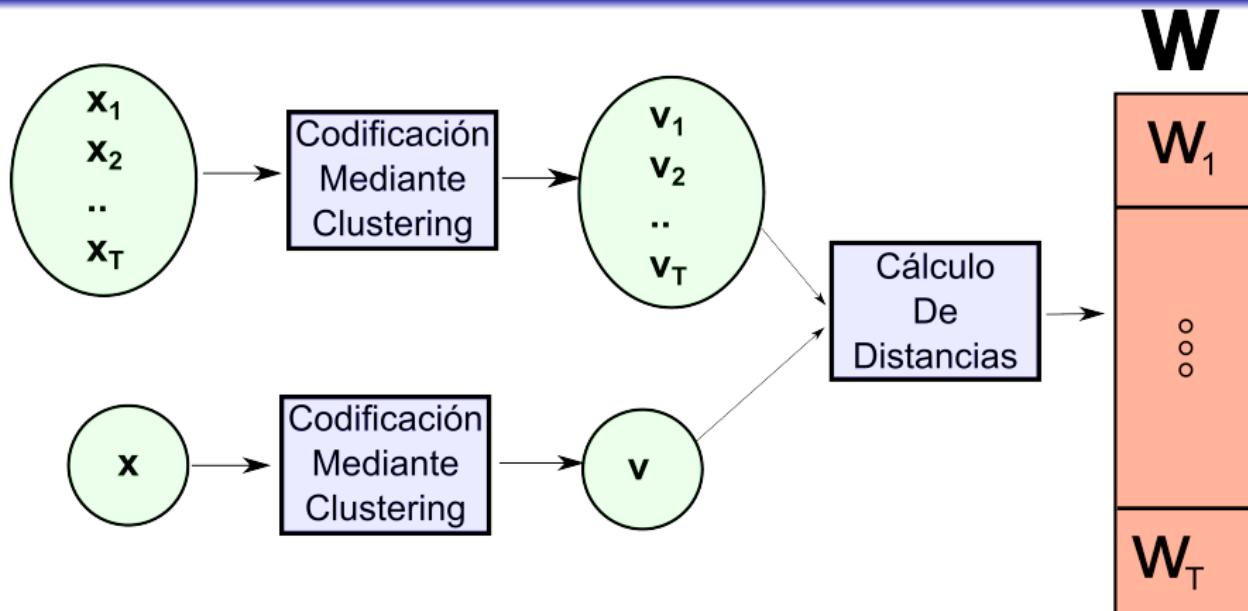
CNC: Funcionamiento de cada red

Entrenamiento capa 3



CNC: Funcionamiento de cada red

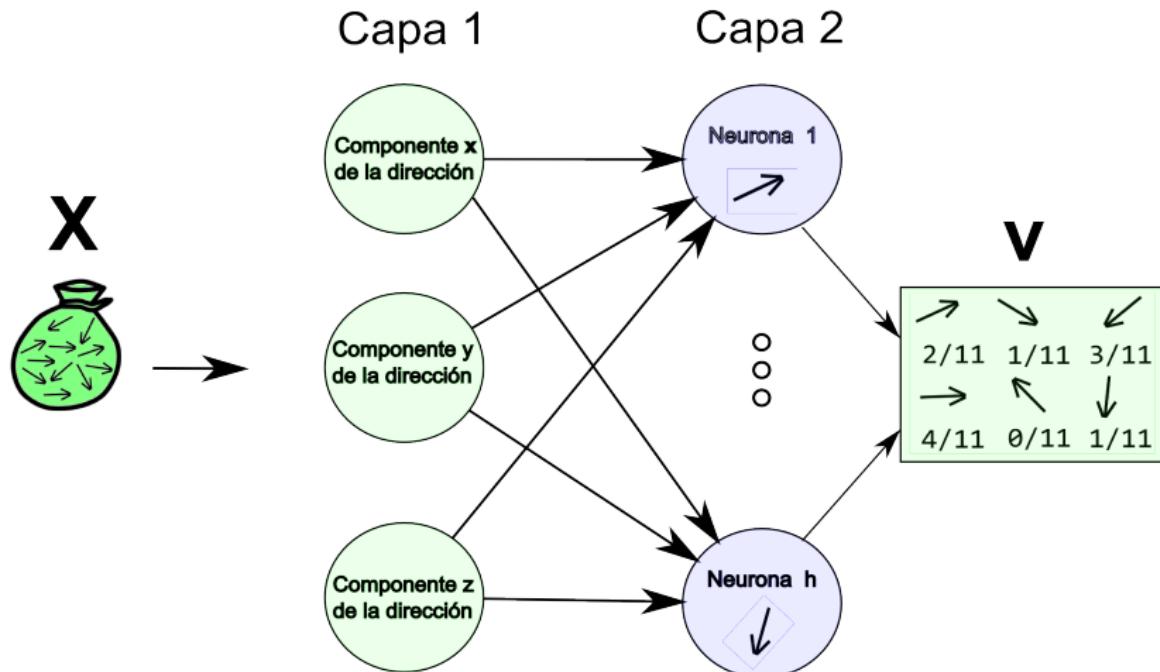
Idea: Asignación de clase a un gesto nuevo



- $T =$ Cantidad de gestos del conjunto de entrenamiento
- $\mathbf{W} =$ Vector de distancias entre \mathbf{v} y \mathbf{v}_i , $i = 1, \dots, T$

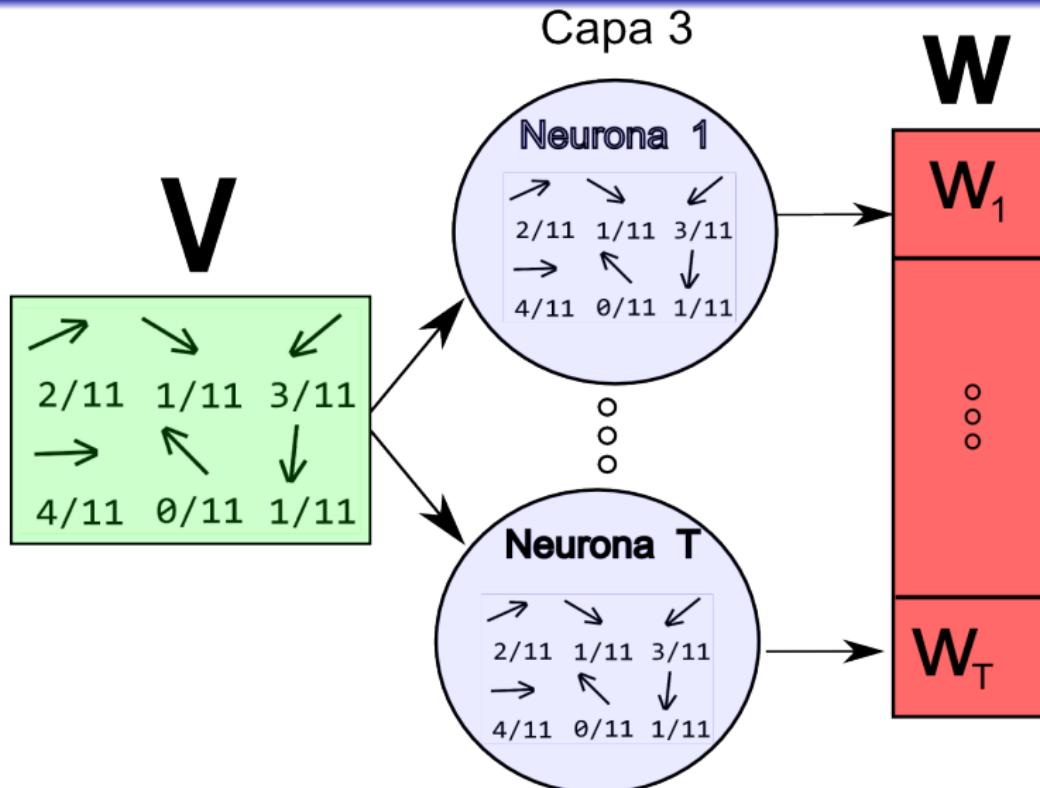
CNC: Funcionamiento de cada red

Funcionamiento capa 2



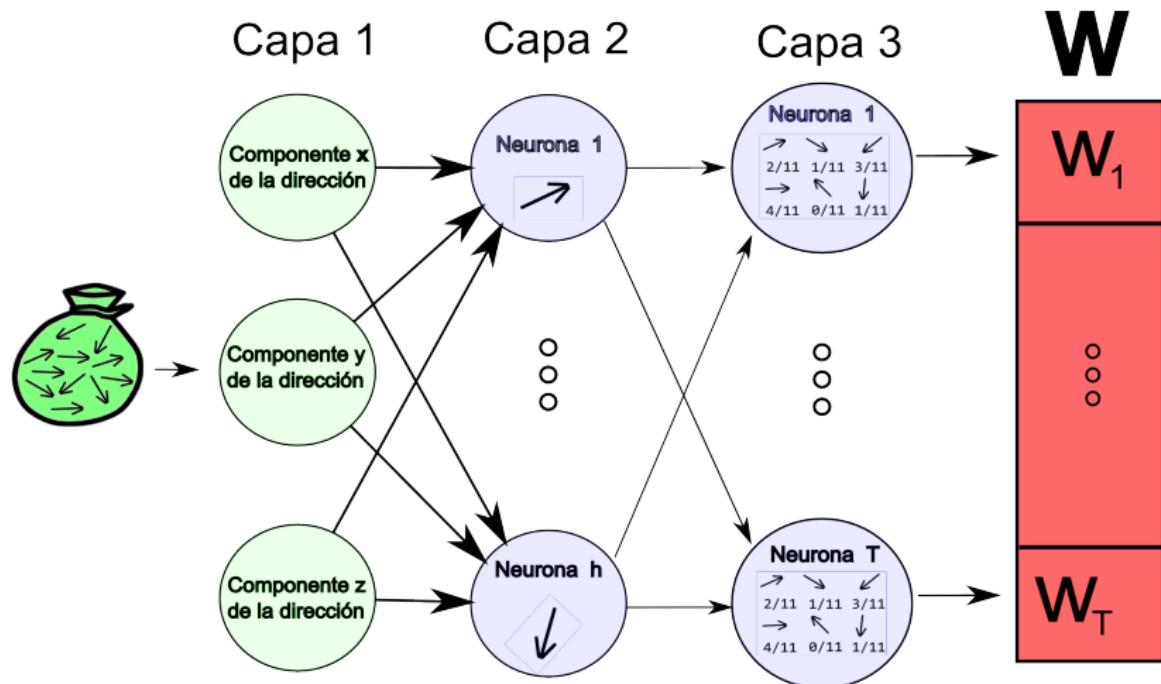
CNC: Funcionamiento de cada red

Funcionamiento capa 3



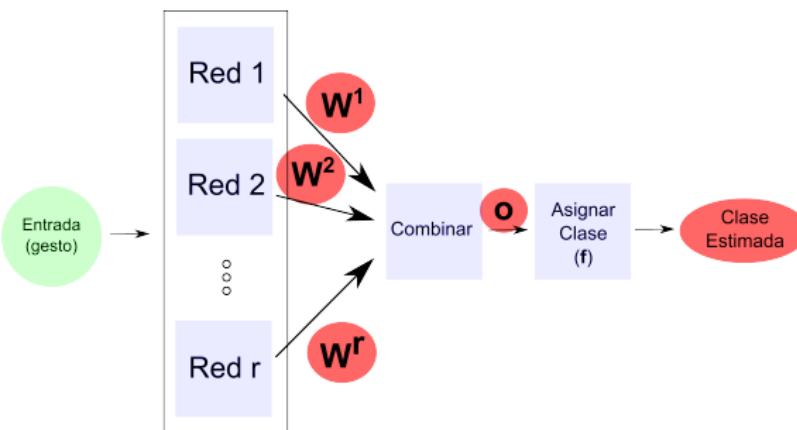
CNC: Funcionamiento de cada red

Estructura Final de cada red



Bagging

Bagging



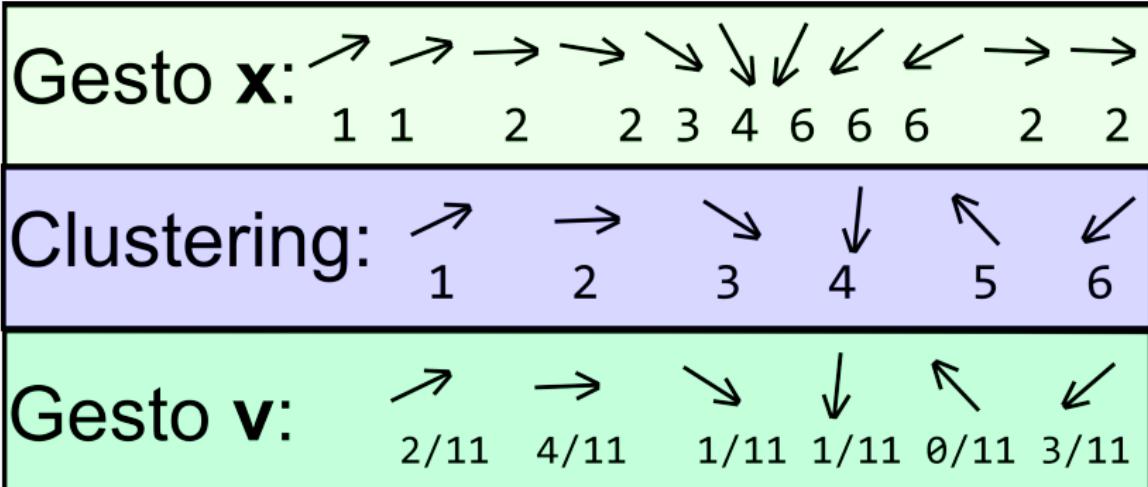
Salida del clasificador para un gesto \mathbf{x}

$$o = \sum_{i=1}^r \mathbf{w}^i$$

$$k = \min_j o_j, j = 1, \dots, T$$

$$f(\mathbf{x}) = clase(k)$$

CNC* = CNC sin re-muestrear los gestos



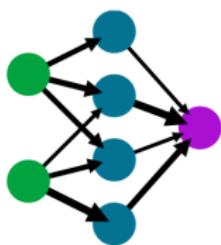
El CNC genera una codificación de tamaño constante

- Tamaño del gesto codificado = h = cantidad de neuronas de la capa 2

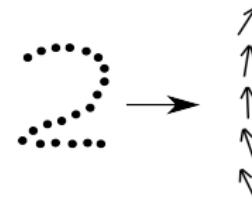
Resultados



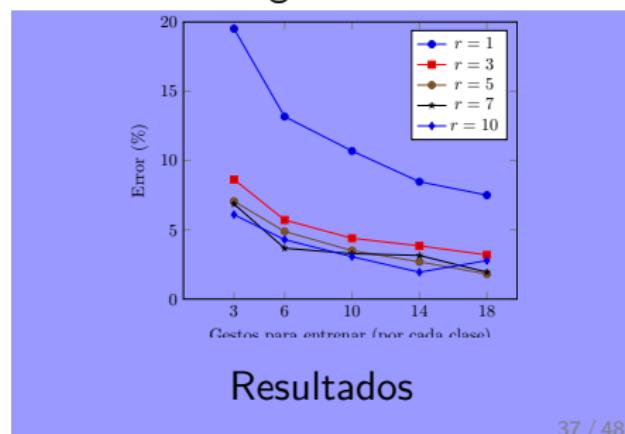
Base de datos de gestos



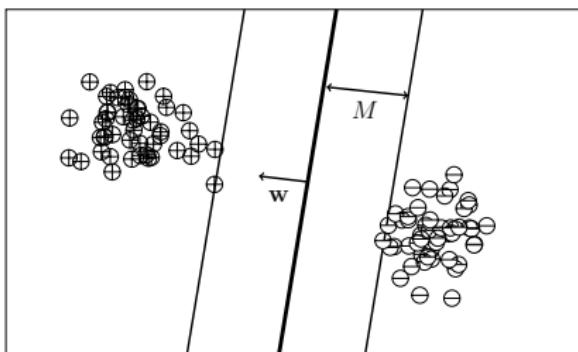
Clasificador Neuronal
Competitivo (CNC)



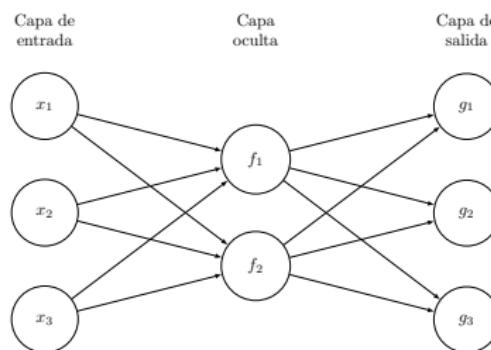
Modelo y representaciones de
gestos



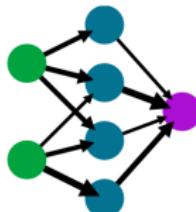
Modelos a comparar



Máquinas de Vectores de Soporte
(SVM)



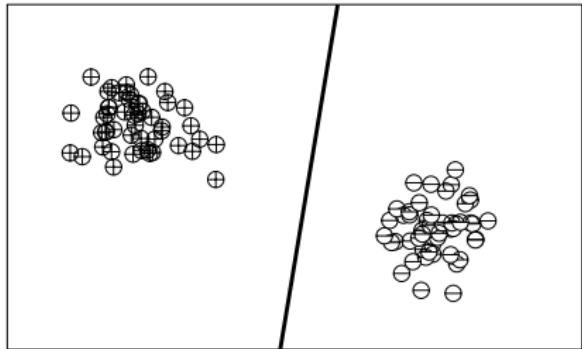
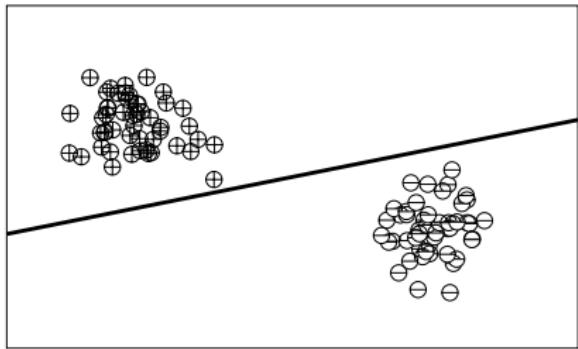
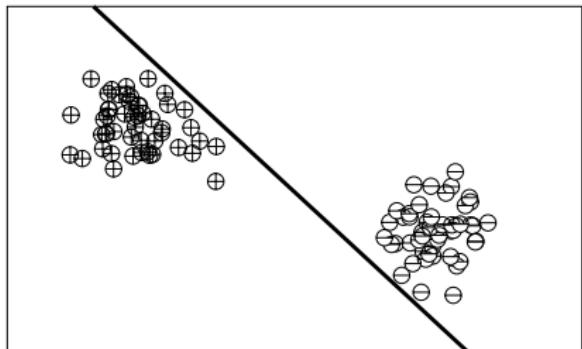
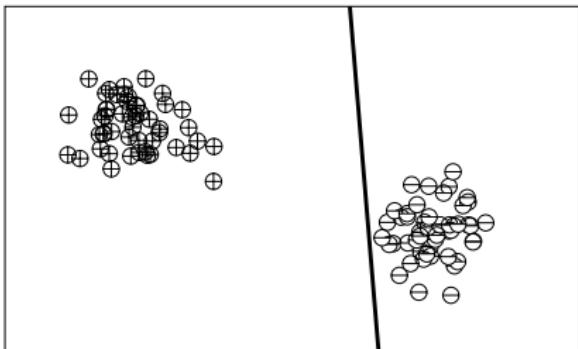
Redes Neuronales Feedforward
(FF)



Clasificador Neuronal Competitivo (CNC)

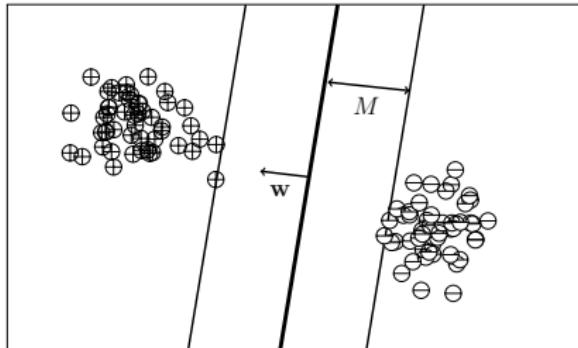
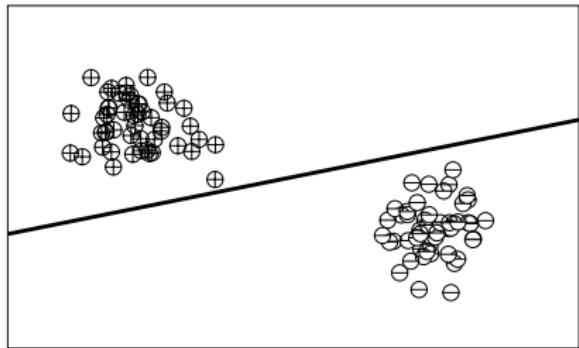
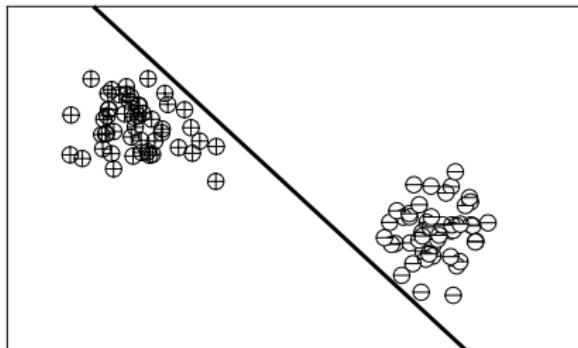
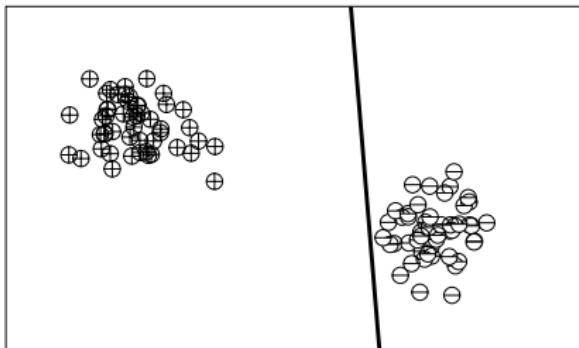
Modelos para comparar

Modelo: Máquinas de Vectores de Soporte (SVM)

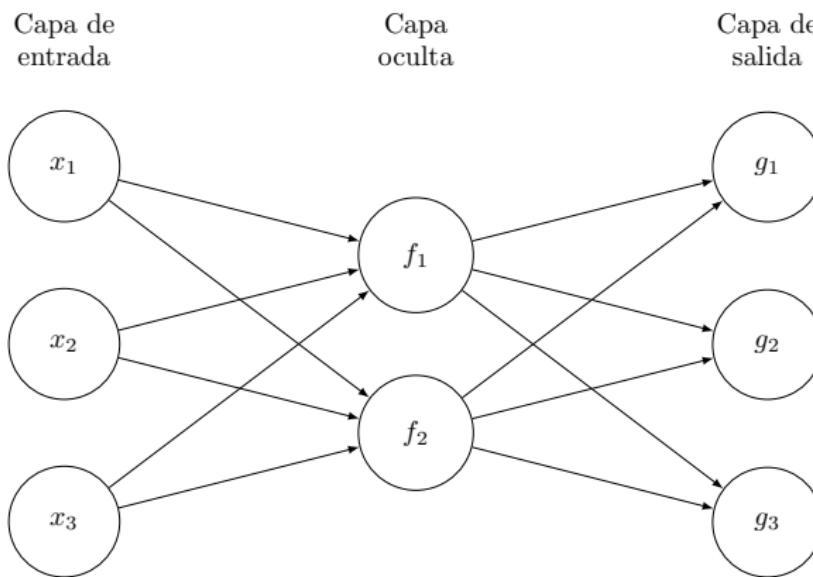


Modelos para comparar

Modelo: Máquinas de Vectores de Soporte (SVM)

Entrenamiento = Separar clases y maximizar el margen M

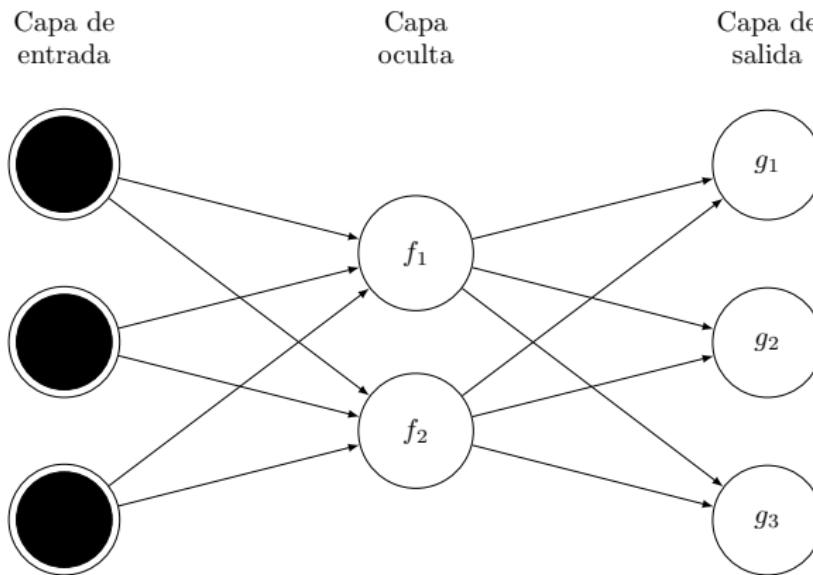
Redes neuronales feedforward (FF)



Grafo de computación

- Ocultas y salida: Dinámica de la red
- Entrenamiento con algoritmo Backpropagation

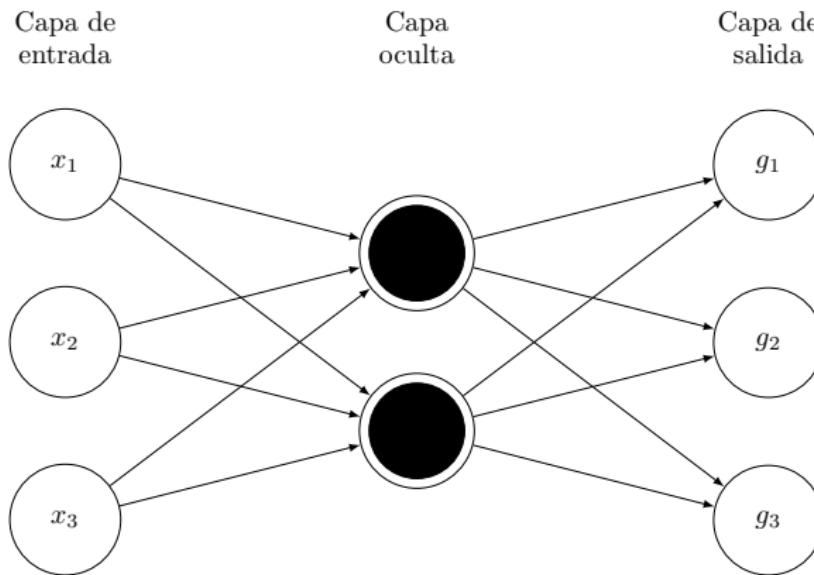
Redes neuronales feedforward (FF)



Grafo de computación

- Ocultas y salida: Dinámica de la red
- Entrenamiento con algoritmo Backpropagation

Redes neuronales feedforward (FF)

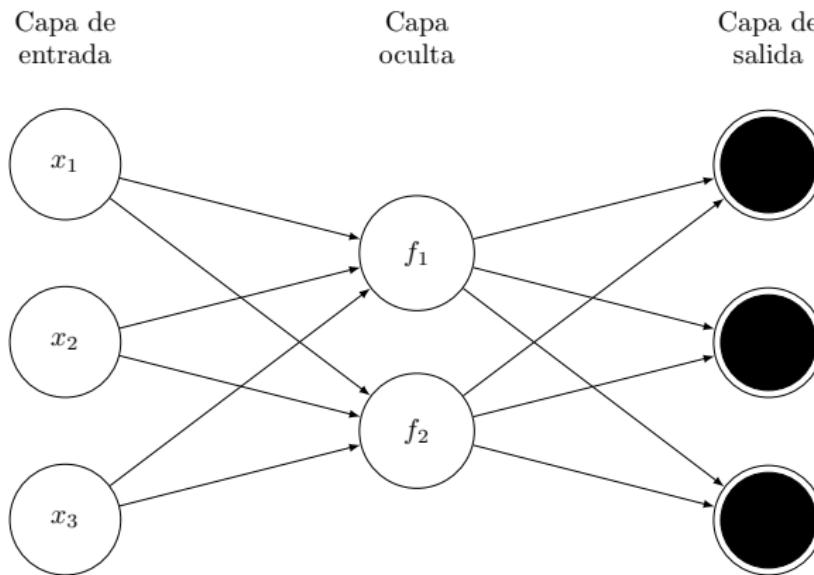


Grafo de computación

- Ocultas y salida: Dinámica de la red
- Entrenamiento con algoritmo Backpropagation

Redes Neuronales Feedforward (FF)

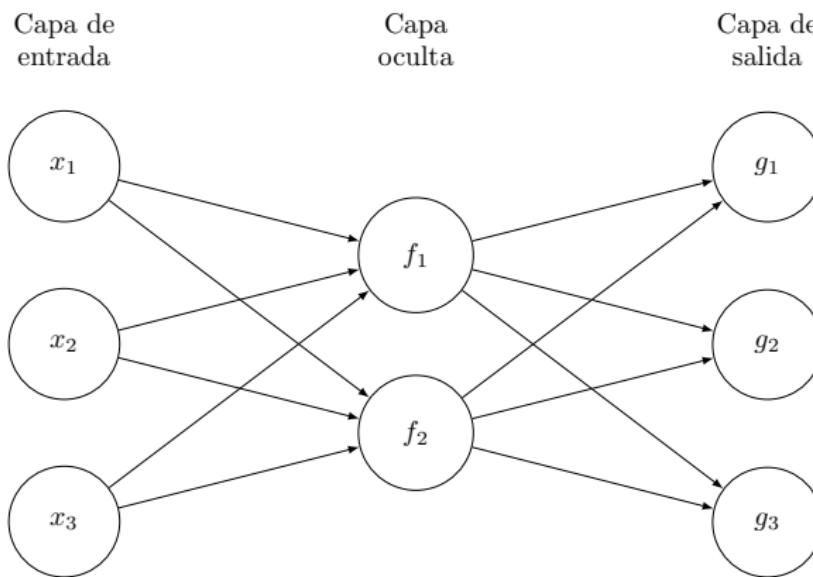
Redes neuronales feedforward (FF)



Grafo de computación

- Ocultas y salida: Dinámica de la red
- Entrenamiento con algoritmo Backpropagation

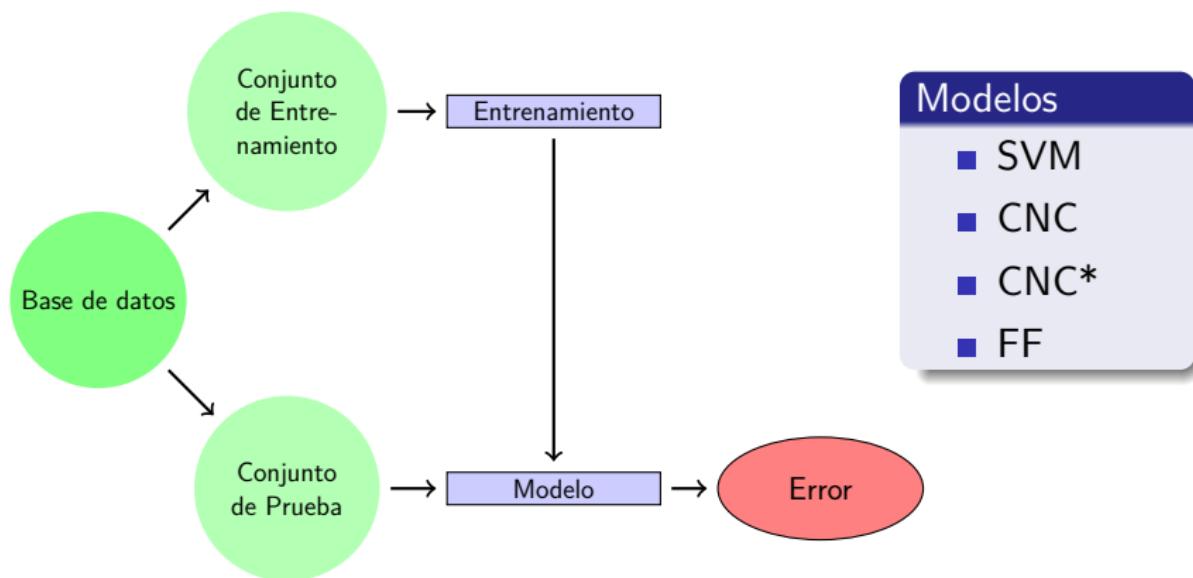
Redes neuronales feedforward (FF)



Grafo de computación

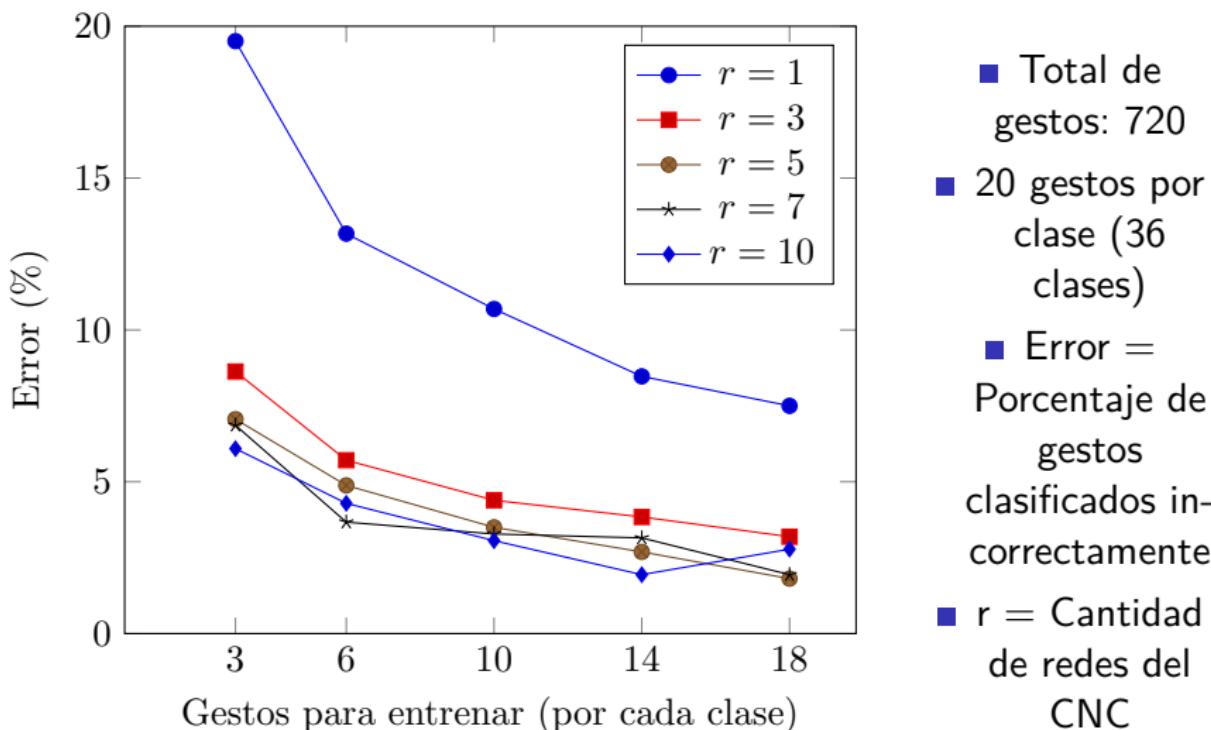
- Ocultas y salida: Dinámica de la red
- Entrenamiento con algoritmo Backpropagation

Setup experimentos

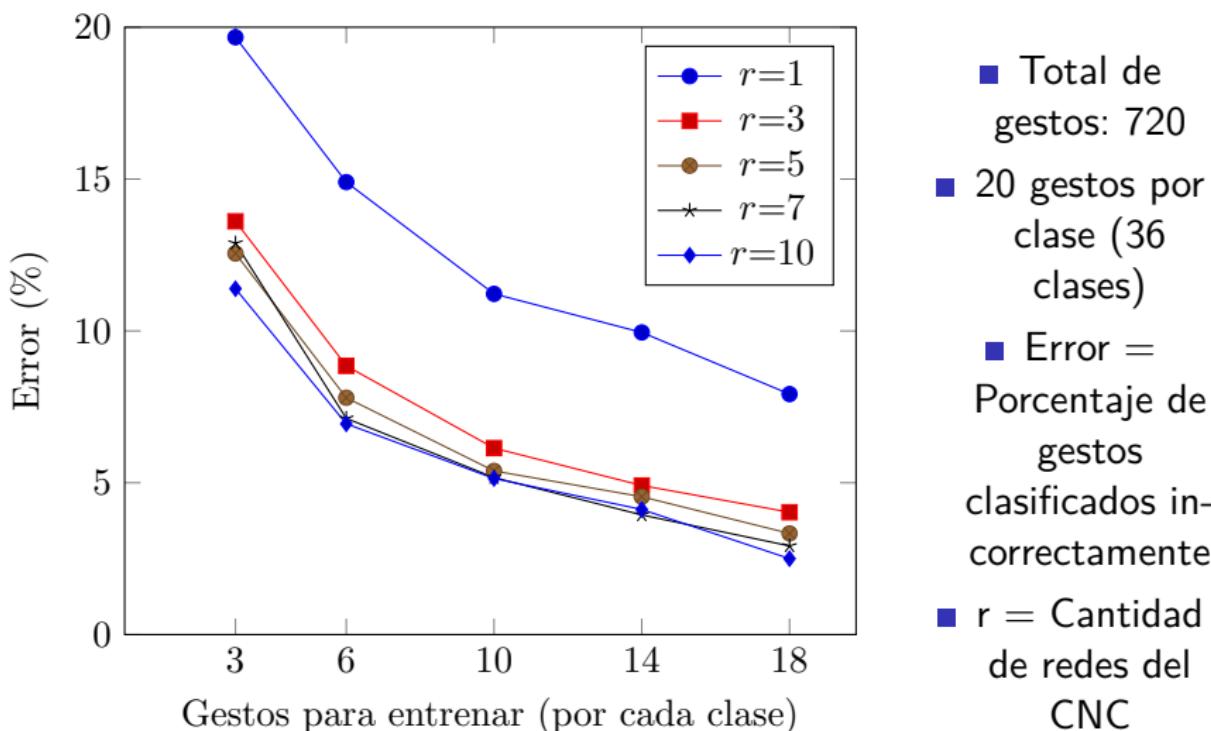


- Pruebas preliminares para determinar parámetros: 30 corridas por cada configuración
- Pruebas finales con mejores parámetros: 500 corridas, mejor configuración

Resultados preliminares: CNC, base de datos LNHG

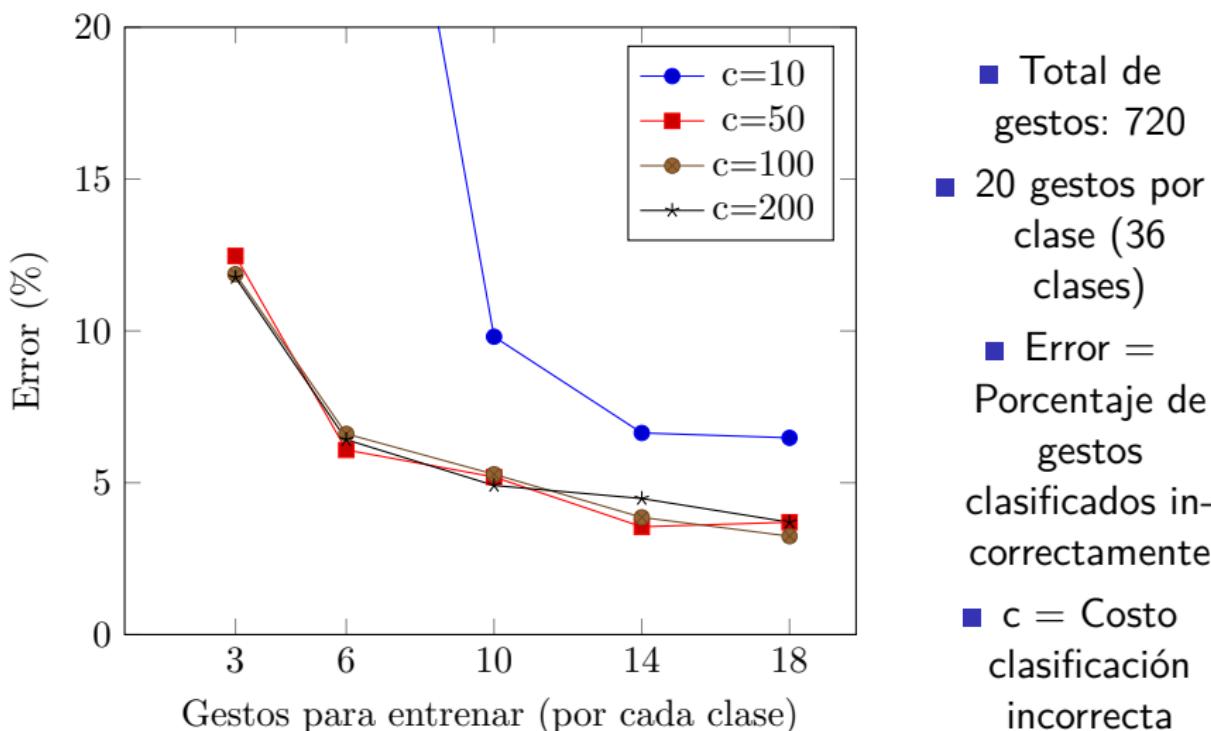


Resultados preliminares: CNC*, base de datos LNHG



Resultados

Resultados preliminares: SVM, base de datos LNHG



Resultados finales: Base de datos LNHG, 500 corridas

Clasificador	Parámetros	Error = $\mu (\sigma)$
SVM	$c = 100,$ $n = 128$	11.76 % (0.080)
CNC	$h = 50,$ $r = 10$ $n = 32$	5.92 % (0.088)
CNC*	$h = 50,$ $r = 10$	11.29 % (0.092)
FF	$h = 100, n =$ 16	18.60 % (0.158)

Ventajas del CNC

- Menos sensible a la cantidad de ejemplares de entrenamiento
- Puede utilizarse sin re-muestreo (CNC*)

Puntos a mejorar

- El CNC pierde toda información de la secuencia de direcciones
→ “Colisiones”: gestos distintos con la misma distribución de direcciones
- Pruebas de tiempo real no satisfactorias
→ Requiere segmentación de buena calidad
- Pocos gestos
→ Las únicas bases de datos de gestos con la mano son de imágenes/video

Trabajo futuro

- Reconocimiento en tiempo real
- “Spotting”: reconocer principio y fin del gesto
- Grabar más ejemplos y clases de gestos
- Utilizar métodos de bagging con todos los algoritmos
- Probar otras características (Coeficientes de DFT normalizados o momentos de Hu).

Fin

Gracias!