

29/03/2023

Hierarchical clustering

Va agrupando samples por distancia (que se puede escoger) hasta tener todos los samples en un unico cluster → Se puede escojer el punto de cluster

FILLING

Agglomerative Hierarchical clustering

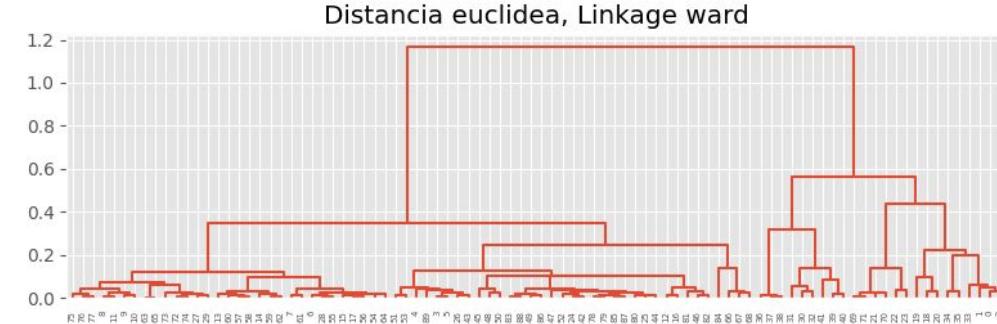
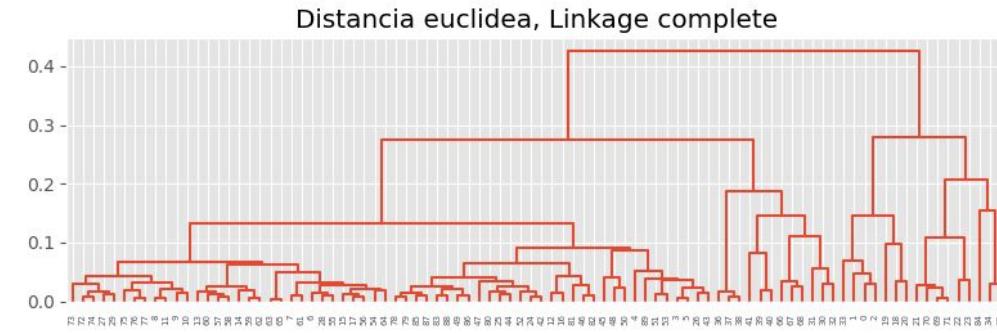
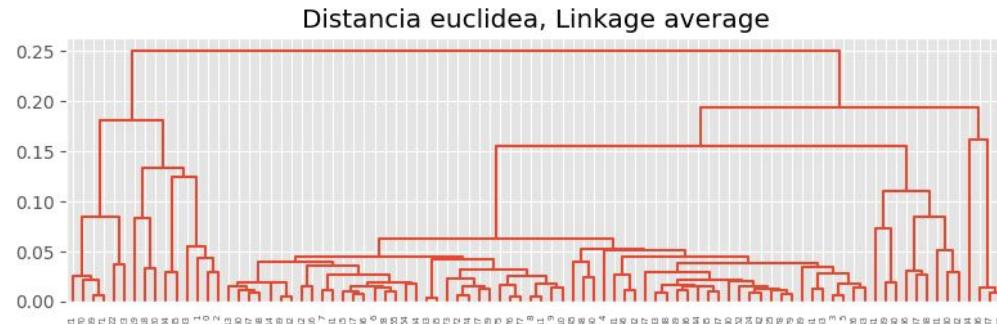
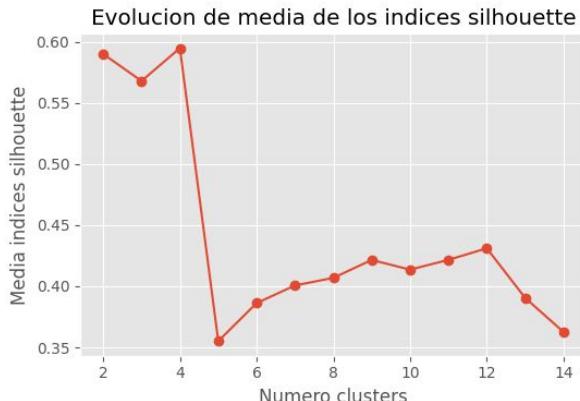
Distancia euclidea

Diferentes linkage (que samples coje para medir distancia)

Nnumero optimo clusters

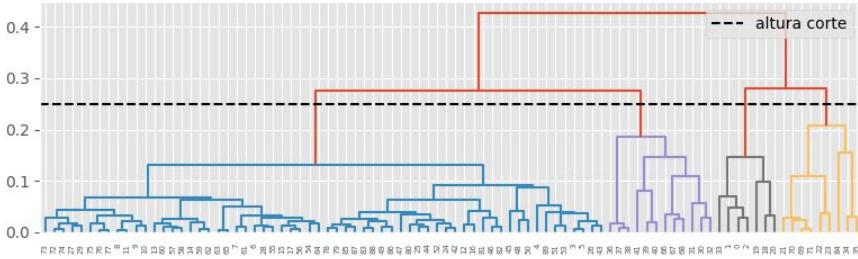
Average silhouette

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

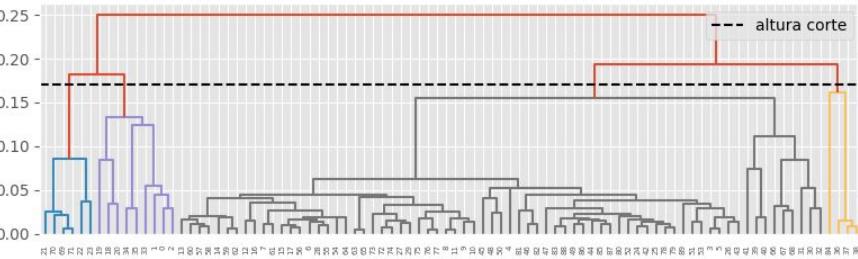


Agglomerative Hierarchical clustering (2x2)

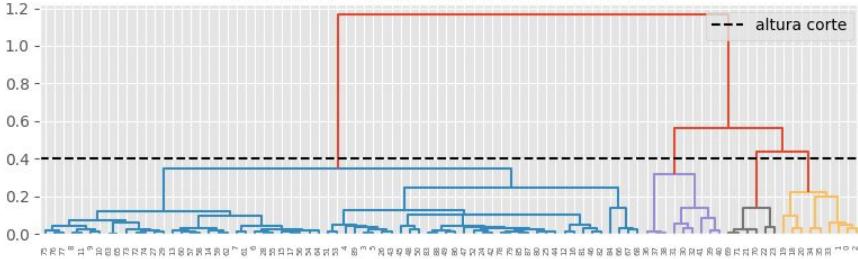
Distancia euclidean, Linkage complete



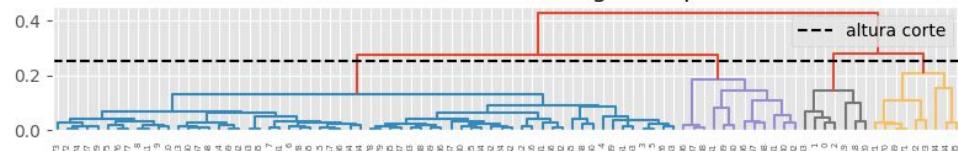
Distancia euclidean, Linkage complete



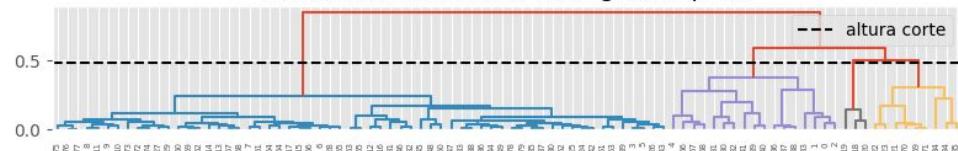
Distancia euclidean, Linkage complete



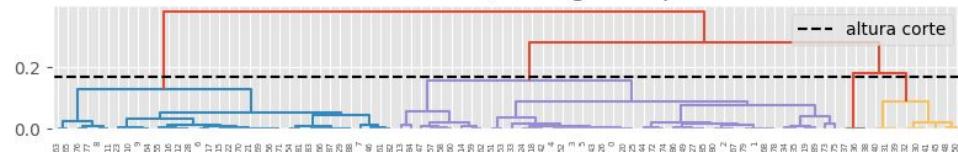
Distancia euclidean, Linkage complete



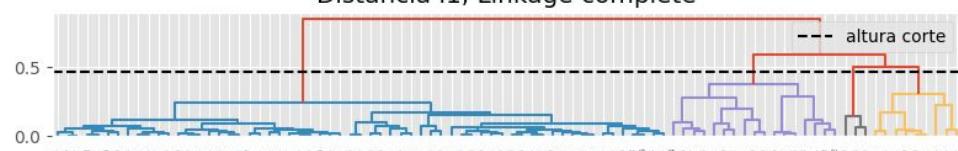
Distancia manhattan, Linkage complete



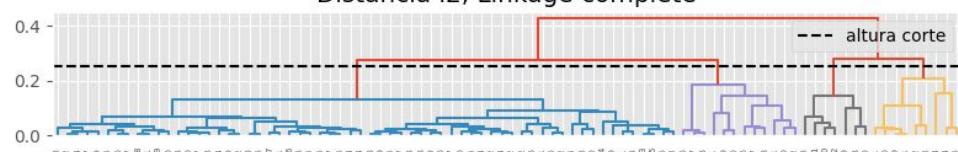
Distancia cosine, Linkage complete



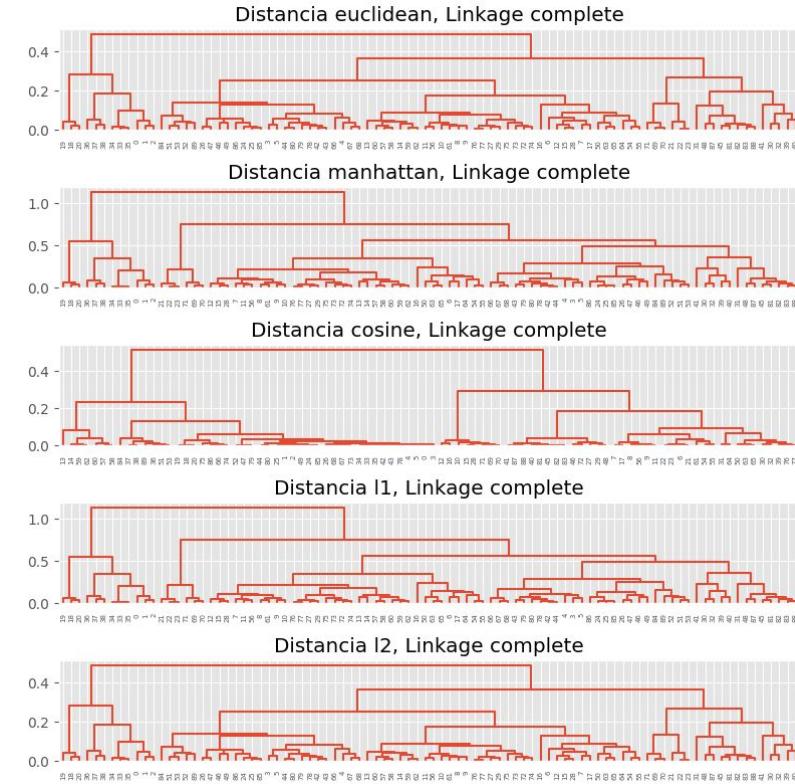
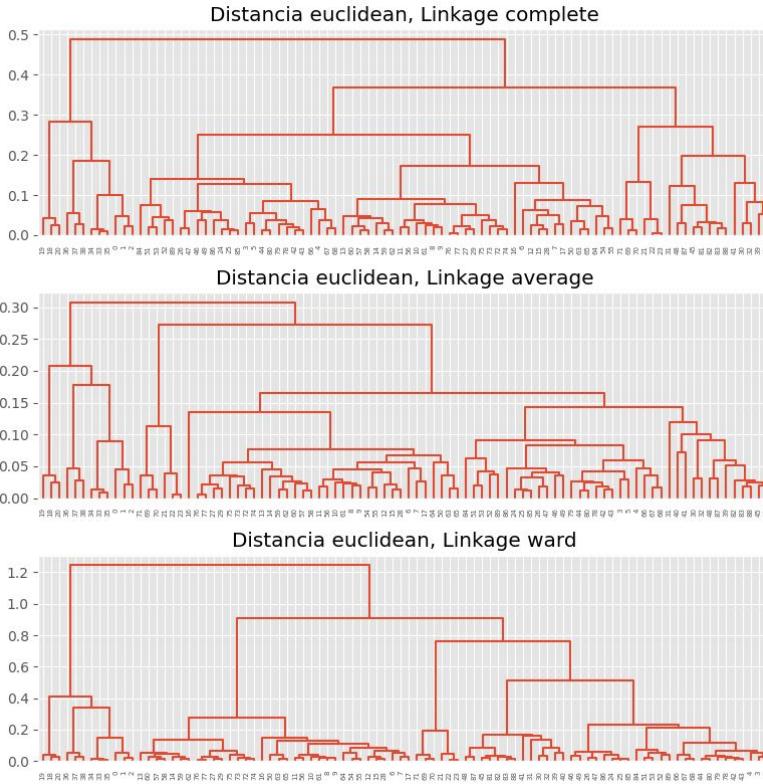
Distancia l1, Linkage complete



Distancia l2, Linkage complete

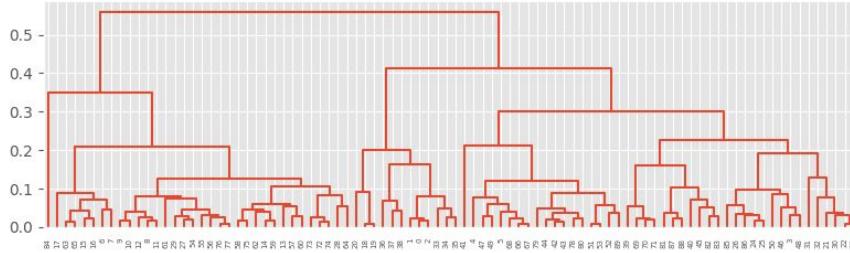


3x3

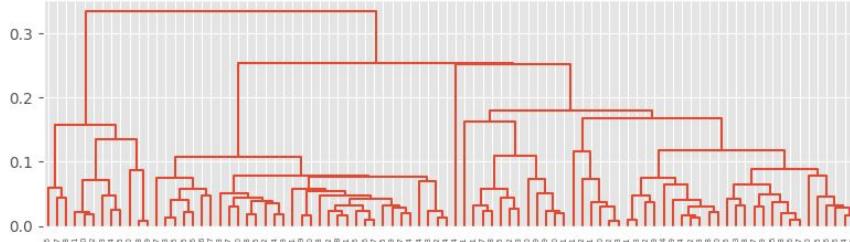


4x4

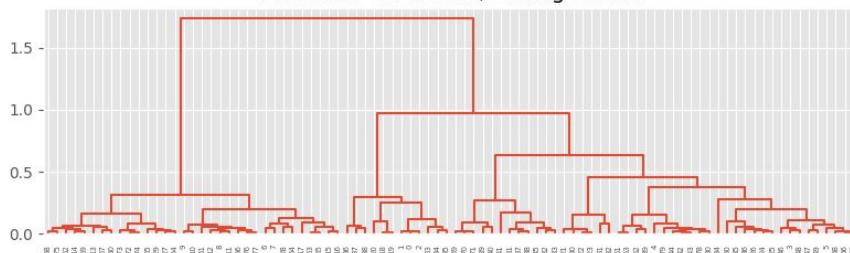
Distancia euclidean, Linkage complete



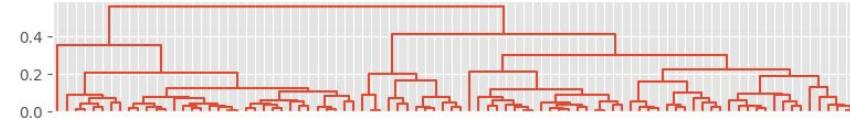
Distancia euclidean, Linkage average



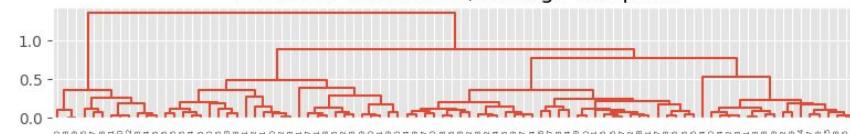
Distancia euclidean, Linkage ward



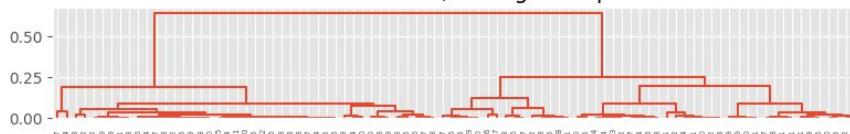
Distancia euclidean, Linkage complete



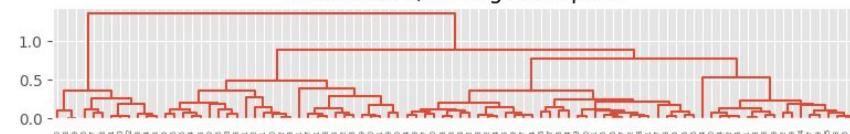
Distancia manhattan, Linkage complete



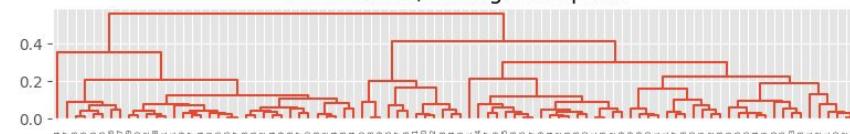
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

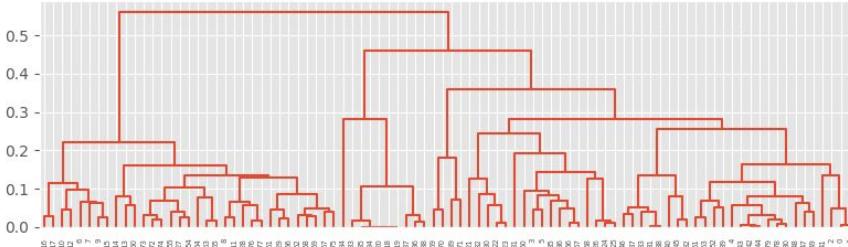


Distancia l2, Linkage complete

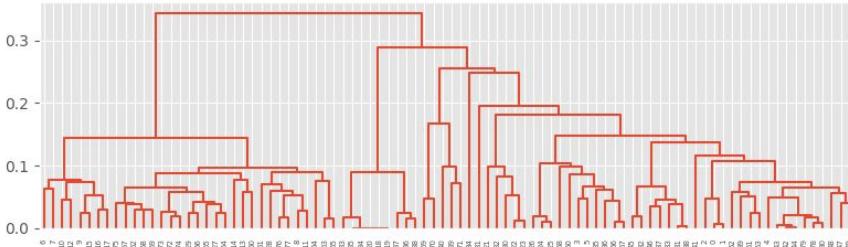


5x5

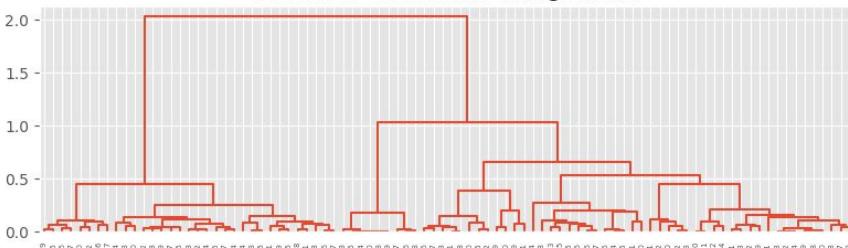
Distancia euclidean, Linkage complete



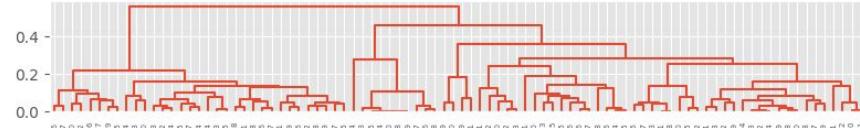
Distancia euclidean, Linkage average



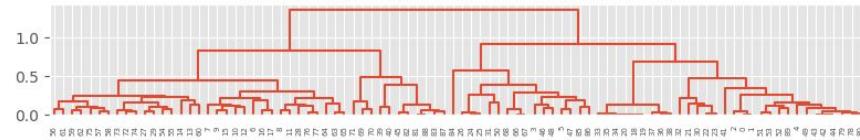
Distancia euclidean, Linkage ward



Distancia euclidean, Linkage complete



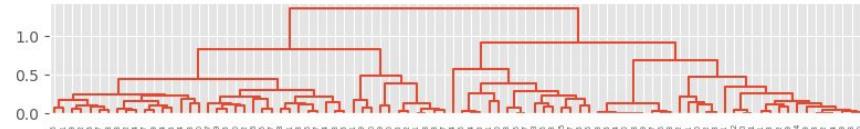
Distancia manhattan, Linkage complete



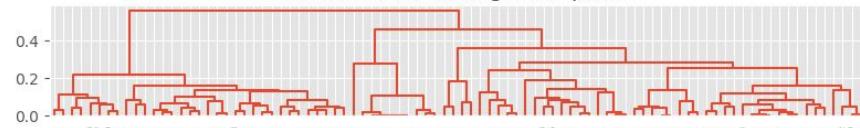
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

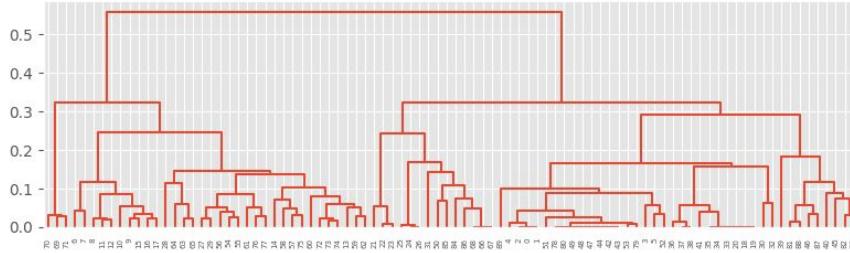


Distancia l2, Linkage complete

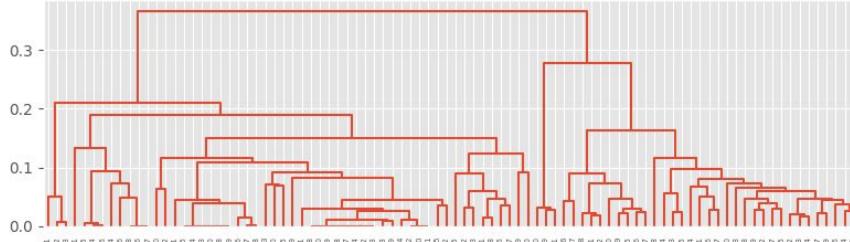


6x6

Distancia euclidean, Linkage complete



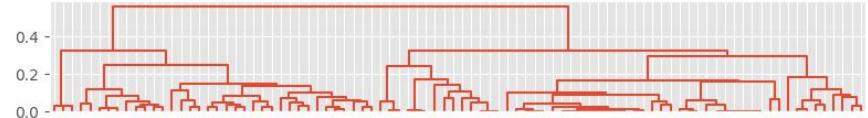
Distancia euclidean, Linkage average



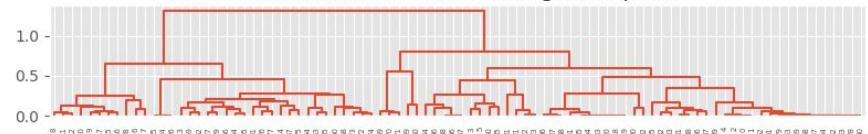
Distancia euclidean, Linkage ward



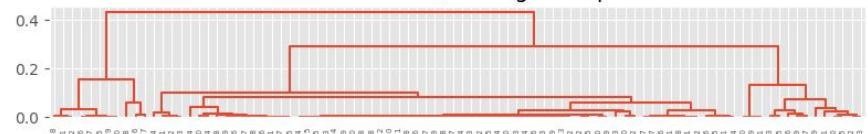
Distancia euclidean, Linkage complete



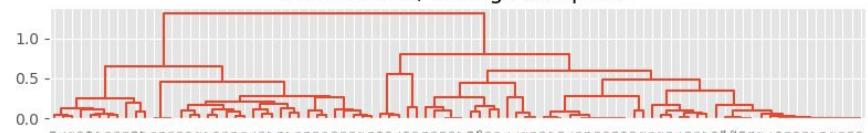
Distancia manhattan, Linkage complete



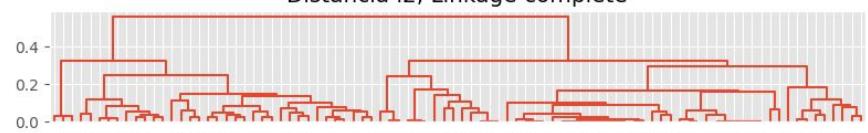
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

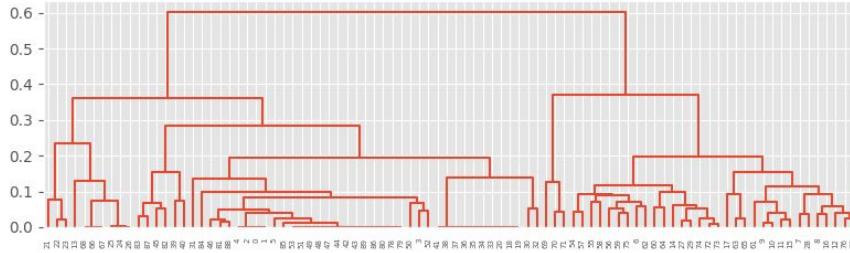


Distancia l2, Linkage complete

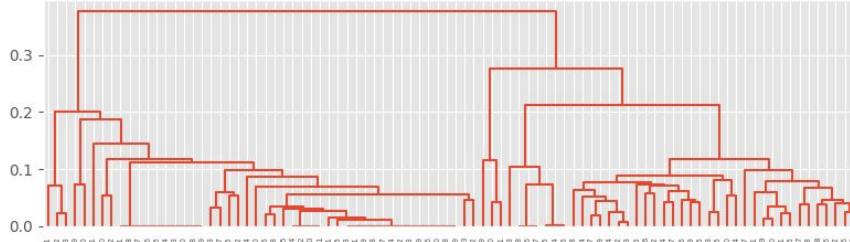


7x7

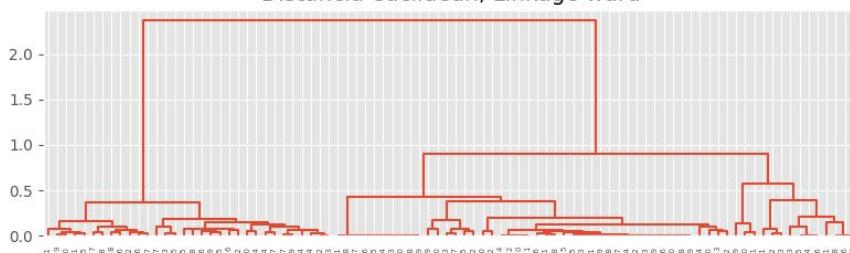
Distancia euclidean, Linkage complete



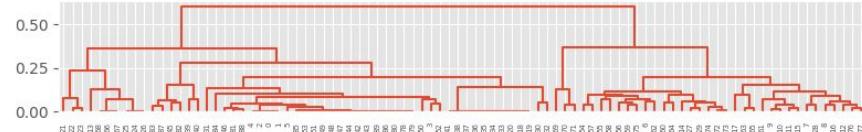
Distancia euclidean, Linkage average



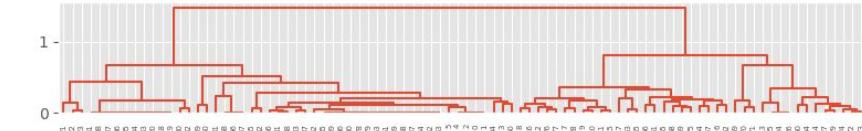
Distancia euclidean, Linkage ward



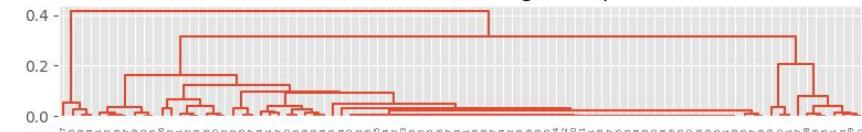
Distancia euclidean, Linkage complete



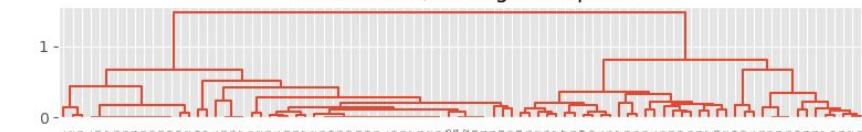
Distancia manhattan, Linkage complete



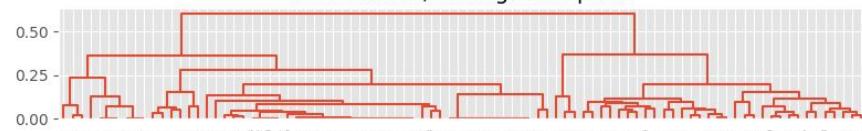
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

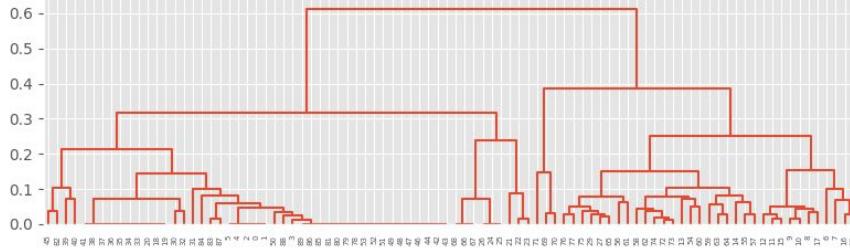


Distancia l2, Linkage complete

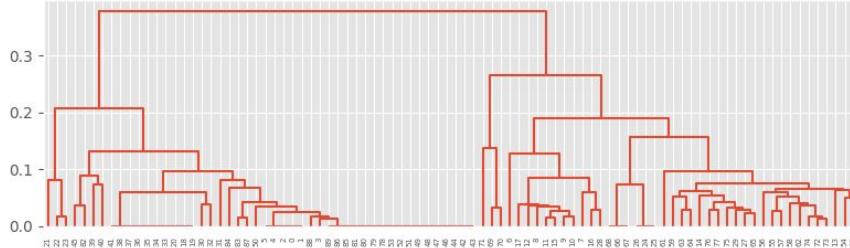


8x8

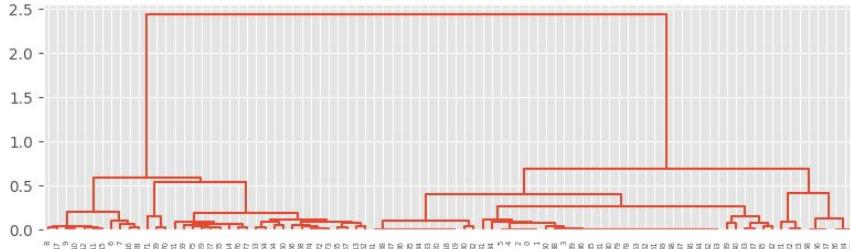
Distancia euclidean, Linkage complete



Distancia euclidean, Linkage average



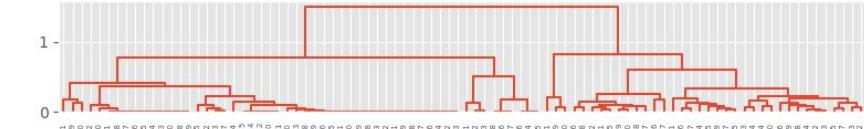
Distancia euclidean, Linkage ward



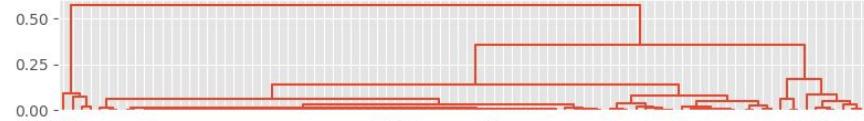
Distancia euclidean, Linkage complete



Distancia manhattan, Linkage complete



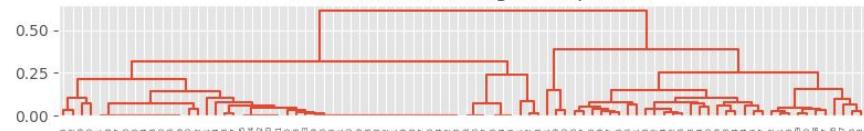
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

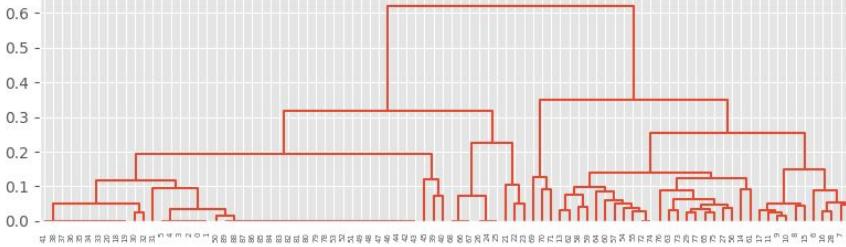


Distancia l2, Linkage complete

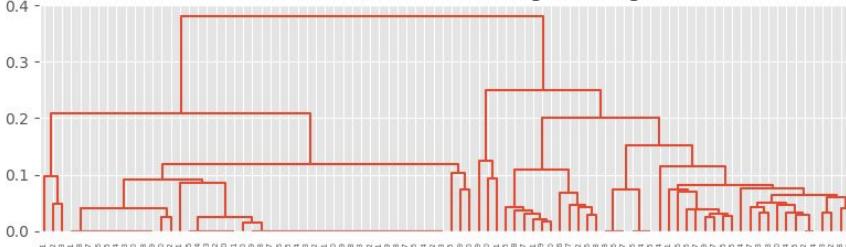


9x9

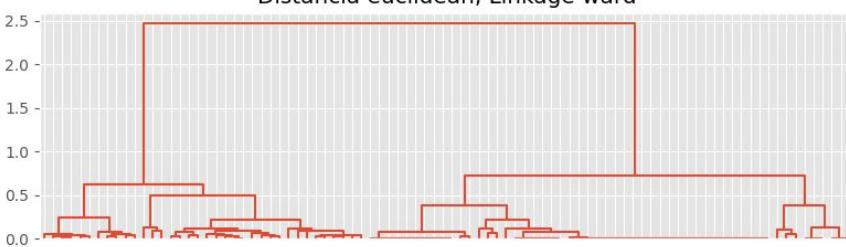
Distancia euclidean, Linkage complete



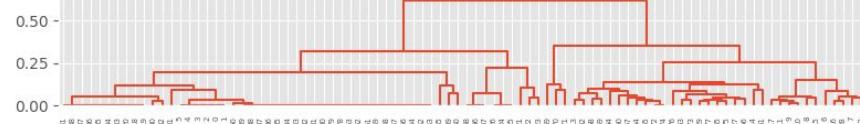
Distancia euclidean, Linkage average



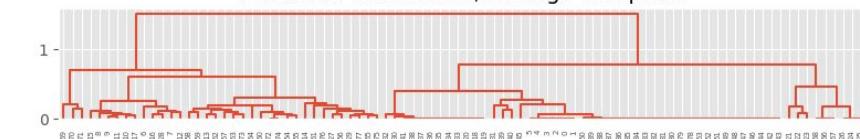
Distancia euclidean, Linkage ward



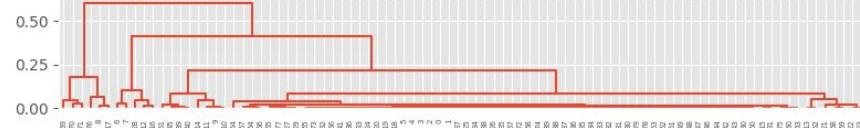
Distancia euclidean, Linkage complete



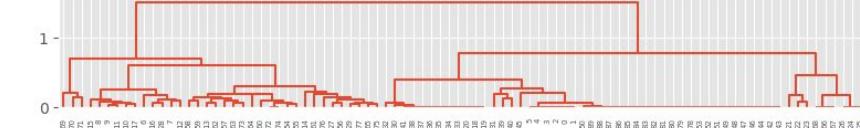
Distancia manhattan, Linkage complete



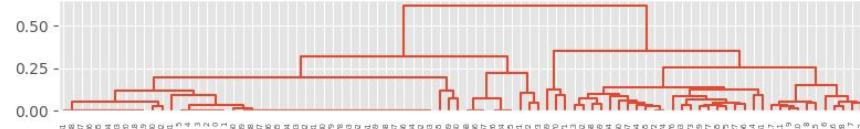
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

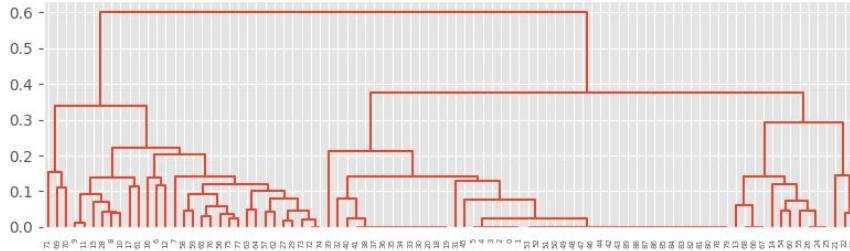


Distancia l2, Linkage complete

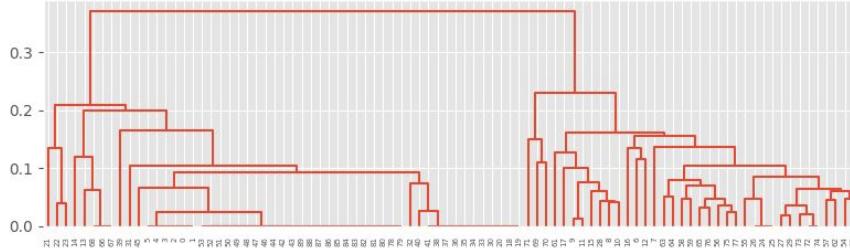


10x10

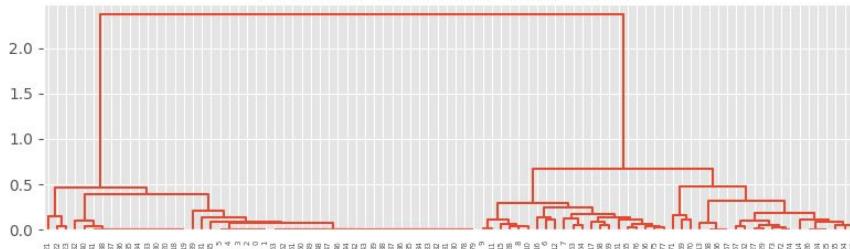
Distancia euclidean, Linkage complete



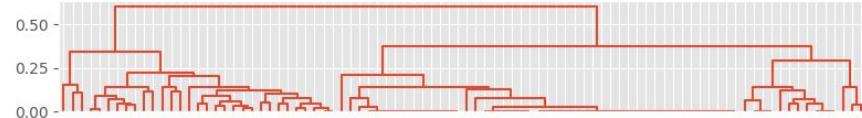
Distancia euclidean, Linkage average



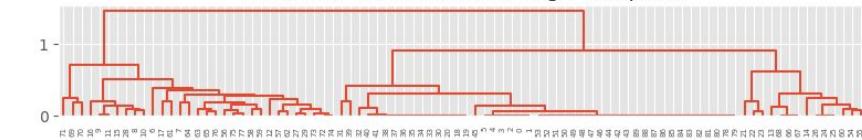
Distancia euclidean, Linkage ward



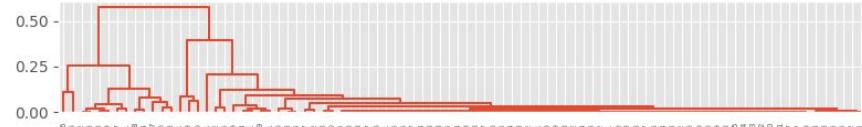
Distancia euclidean, Linkage complete



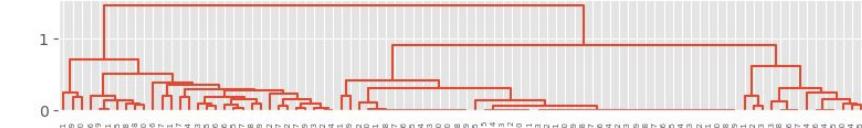
Distancia manhattan, Linkage complete



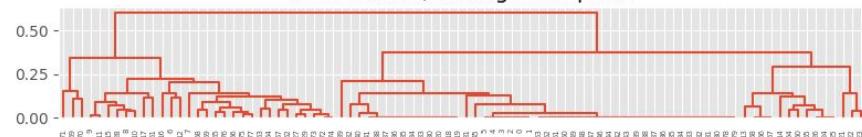
Distancia cosine, Linkage complete



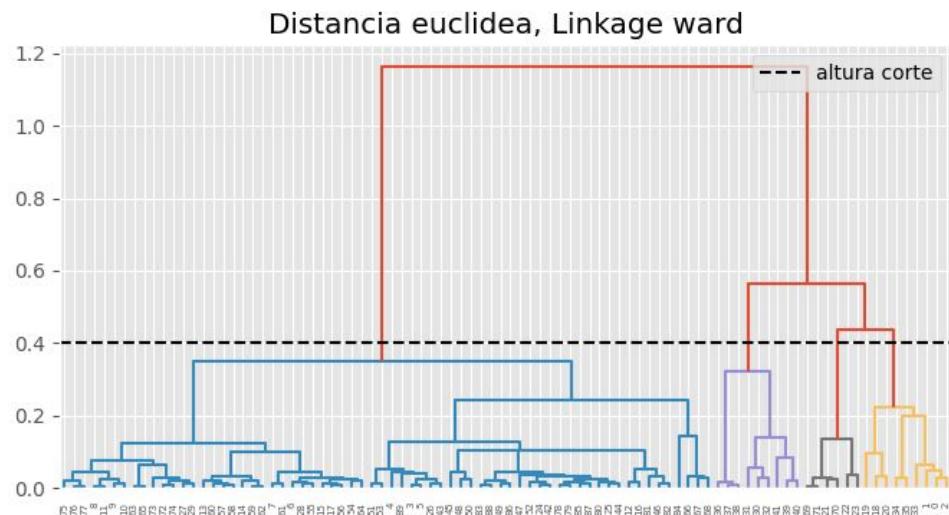
Distancia l1, Linkage complete



Distancia l2, Linkage complete



Agglomerative Hierarchical clustering



Success rates - aggl. hier. clustering, 4 clases

- Euclidean, Ward

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	100	100	95.429	60.714	60.714	75
Success_B	20	26.667	51.111	55.556	73.333	75.556	75.556	75.556	24.444
Success_C	54.545	90.909	0	0	9.091	0.091	0	0	100
Success_D	100	100	50	50	50	50	0	0	50
TOTAL	54.444	62.222	60	62.222	72.222	72.222	56.667	56.667	51.111

- Euclidean, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	100	100	3.571	100	100	82.143
Success_B	13.333	20	62.222	68.889	71.111	75.556	75.556	75.556	13.333
Success_C	0	72.727	0	9.091	27.273	0	0	0	100
Success_D	100	0	0	50	50	0	50	50	50
TOTAL	44.444	50	62.222	70	73.333	38.889	72.222	72.222	47.778

- Euclidean, average

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	0	100	0	100	96.429	100	100	92.857
Success_B	20	73.333	26.667	68.889	88.889	75.556	75.556	75.556	82.222
Success_C	9.091	0	90.909	0	0	0	0	0	0
Success_D	100	100	0	0	50	0	50	50	50
TOTAL	48.889	43.333	55.556	34.444	78.889	67.778	72.222	72.222	73.333

Success rates - aggl. hier. clustering

- Manhattan, Complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	100	60.714	85.714	100	100	82.143
Success_B	24.444	62.222	26.667	11.111	82.222	62.222	75.556	75.556	13.333
Success_C	9.091	0	81.818	36.364	18.182	0	0	0	100
Success_D	0	0	0	50	0	50	50	50	50
TOTAL	44.444	62.222	54.444	44.444	62.222	61.111	72.222	72.222	47.778

- Cosine, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	71.429	53.571	7.143	35.714	17.857	14.286	14.286	17.857	28.571
Success_B	75.556	80	73.333	75.556	93.333	100	100	100	100
Success_C	81.818	0	0	36.364	0	0	0	0	0
Success_D	0	50	50	0	0	50	50	50	0
TOTAL	70	60	42.222	53.333	52.222	57.778	57.778	58.889	58.889

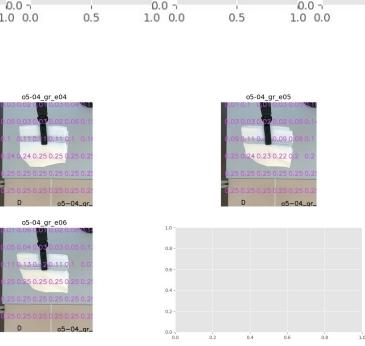
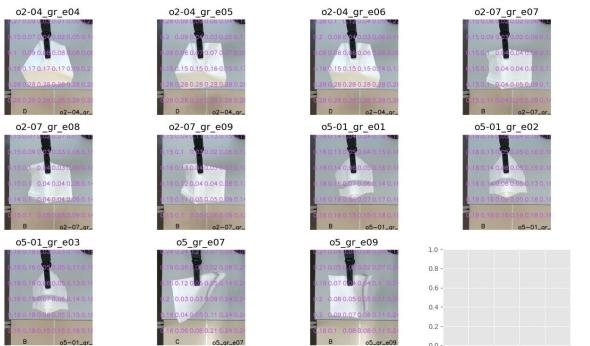
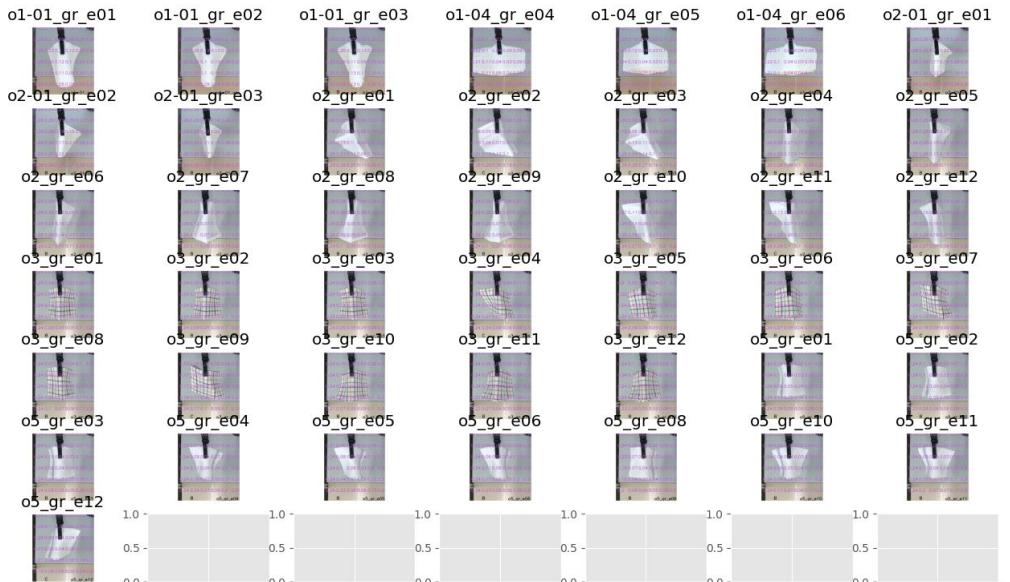
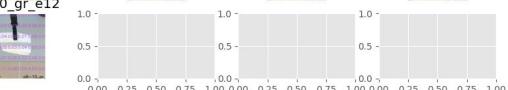
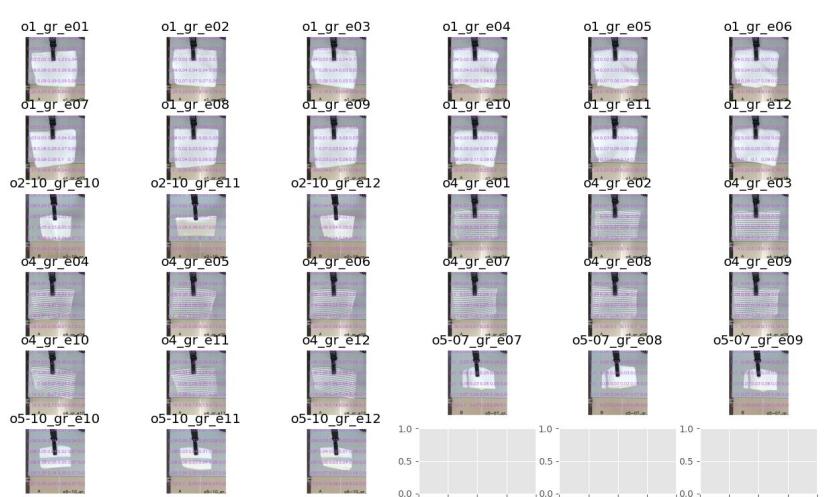
- L1, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	100	60.714	85.714	100	100	82.143
Success_B	24.444	62.222	26.667	11.111	82.222	62.222	75.556	75.556	13.333
Success_C	9.091	0	81.818	36.364	18.182	0	0	0	100
Success_D	0	0	0	50	0	50	50	50	50
TOTAL	44.444	62.222	54.444	44.444	62.222	61.111	72.222	72.222	47.778

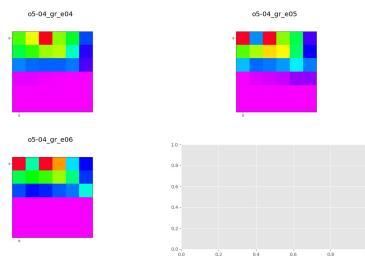
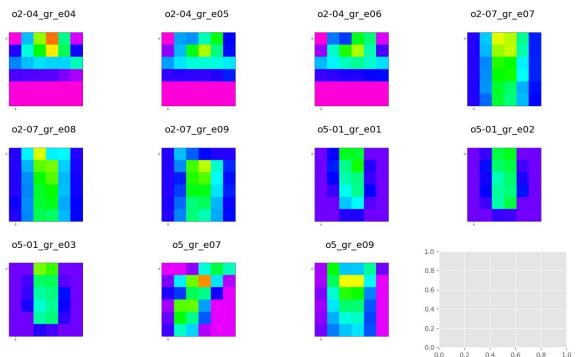
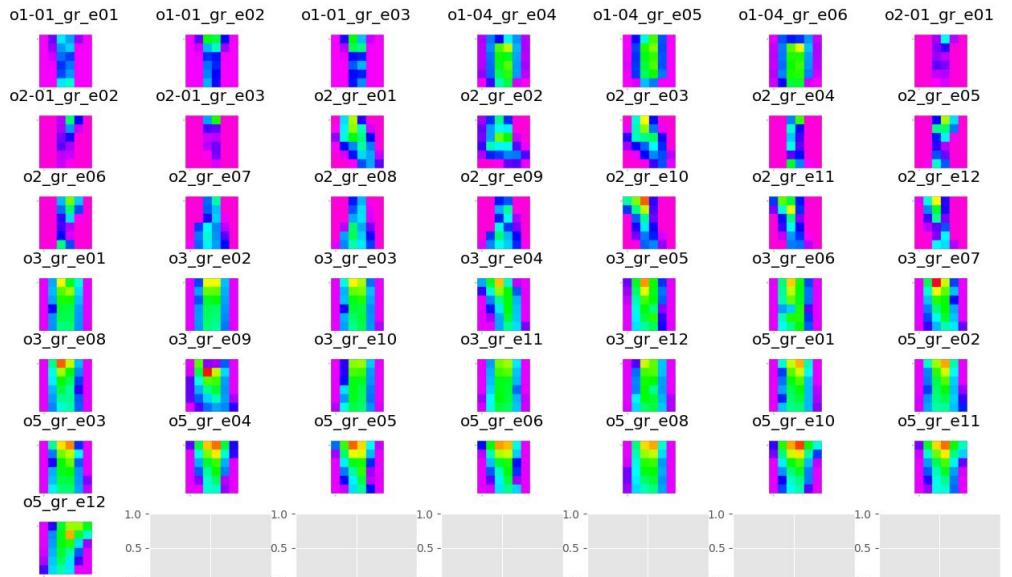
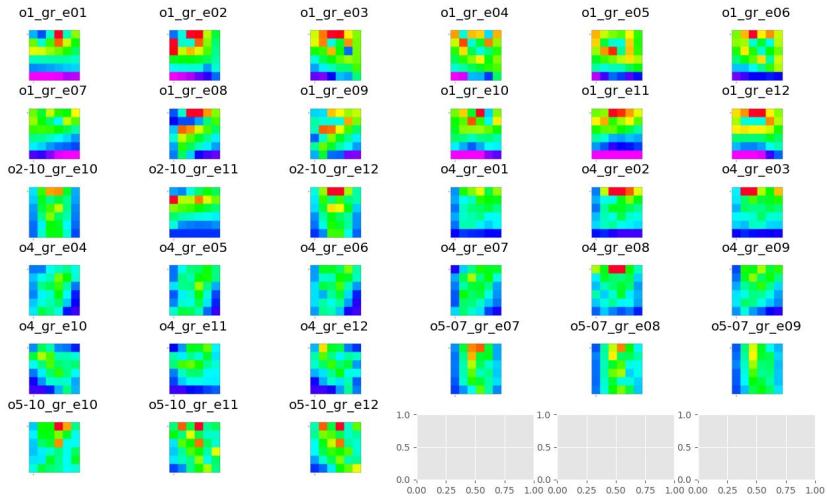
- L2, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	100	100	3.571	100	100	82.143
Success_B	13.333	20	62.222	68.889	71.111	75.556	75.556	75.556	13.333
Success_C	0	72.727	0	9.091	27.273	0	0	0	100
Success_D	100	0	0	50	50	0	50	50	50
TOTAL	44.444	50	62.222	70	73.333	38.889	72.222	72.222	47.778

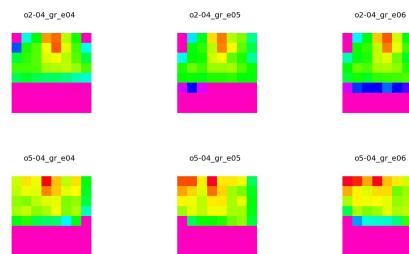
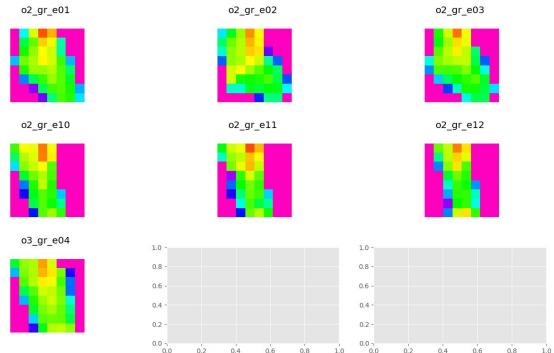
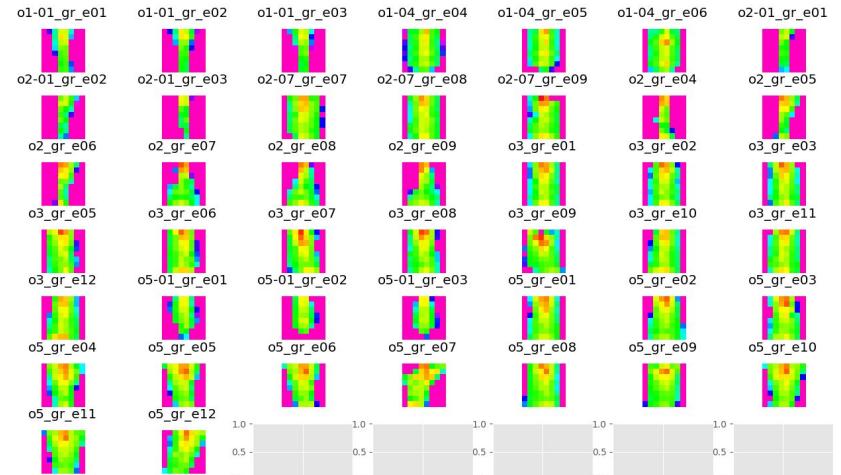
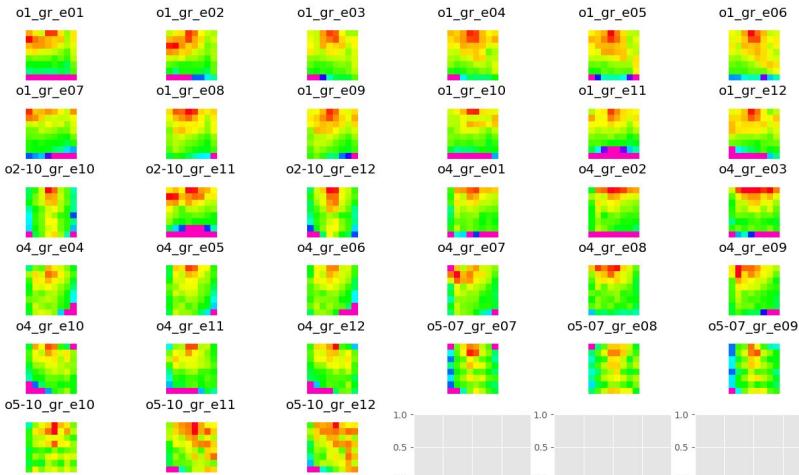
Euclidean, ward, 6x6



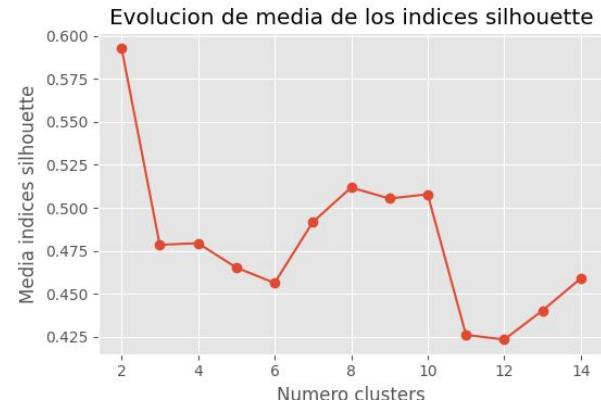
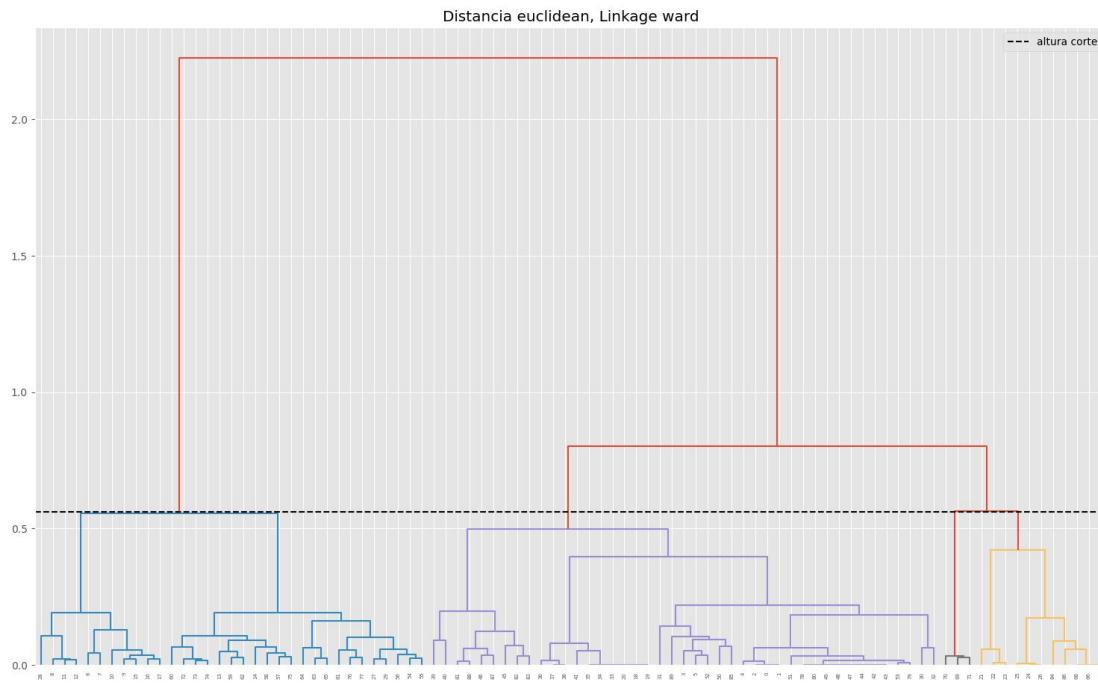
Euclidean, ward, 6x6



not filling peso 8x8



Euclidean, ward, 6x6



NOT FILLING

Success rates - aggl. hier. clustering, 4 clases

Clase D?

- Euclidean, Ward

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	0	100	85.714	39.286	3.571	100	100	25
Success_B	80	71.111	26.667	57.778	80	84.444	11.111	93.333	93.333
Success_C	0	0	0	54.545	0	9.091	27.273	36.364	0
Success_D	83.333	100	100	0	0	0	0	0	0
TOTAL	45.556	42.222	51.111	62.222	52.222	44.444	40	82.222	54.444

- Euclidean, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	0	100	100	96.429
Success_B	95.556	15.556	24.444	55.556	68.889	82.222	86.667	93.333	93.333
Success_C	0	0	9.091	45.455	27.273	9.091	36.364	9.091	9.091
Success_D	0	100	100	0	83.333	0	33.333	0	0
TOTAL	47.778	45.556	51.111	64.444	74.444	42.222	81.111	78.889	77.778

- Euclidean, average

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	14.286	60.714	96.429	100	89.286	25
Success_B	80	17.778	15.556	57.778	68.889	71.111	8.889	93.333	93.333
Success_C	0	0	45.455	45.455	54.545	36.364	27.273	0	0
Success_D	33.333	100	0	0	0	0	33.333	33.333	0
TOTAL	42.222	46.667	44.444	38.889	60	70	41.111	76.667	54.444

Success rates - aggl. hier. clustering - 4x4

- Manhattan, Complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	0	0	96.429
Success_B	82.222	15.556	11.111	48.889	55.556	71.111	86.667	93.333	93.333
Success_C	9.091	0	0	45.455	36.364	36.364	0	0	18.182
Success_D	0	100	0	0	0	0	33.333	33.333	50
TOTAL	42.222	45.556	36.667	61.111	63.333	71.111	45.556	48.889	82.222

- Cosine, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	35.714	67.857	85.714	64.286	67.857	64.286	10.714	7.143	46.429
Success_B	75.556	17.778	46.667	57.778	62.222	75.556	91.111	93.333	93.333
Success_C	0	9.091	18.182	0	9.091	9.091	18.182	0	0
Success_D	0	0	0	0	0	50	33.333	0	16.667
TOTAL	48.889	31.111	52.222	48.889	53.333	62.222	53.333	48.889	62.222

- L1, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	0	0	96.429
Success_B	82.222	15.556	11.111	48.889	55.556	71.111	86.667	93.333	93.333
Success_C	9.091	0	0	45.455	36.364	36.364	0	0	18.182
Success_D	0	100	0	0	0	0	33.333	33.333	50
TOTAL	42.222	45.556	36.667	61.111	63.333	71.111	45.556	48.889	82.222

- L2, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	96.429	100	96.429
Success_B	80	15.556	11.111	40	66.667	71.111	86.667	93.333	93.333
Success_C	0	0	0	18.182	9.091	27.273	36.364	9.091	18.182
Success_D	50	100	0	0	0	0	16.667	0	50
TOTAL	43.333	45.556	36.667	53.333	65.556	70	78.889	78.889	82.222

Success rates - aggl. hier. clustering, 3 clases

Clase D?

- Euclidean, Ward

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	100	0	96.429
Success_B	95.556	15.556	24.444	57.778	13.333	73.333	86.667	93.333	93.333
Success_C	0	0	54.545	18.182	18.182	27.273	36.364	0	9.091
Success_D	0	0	0	0	0	0	0	0	0
TOTAL	47.778	38.889	50	62.222	40	71.111	78.889	46.667	77.778

- Euclidean, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	96.429	100	96.429
Success_B	80	15.556	11.111	40	66.667	71.111	86.667	93.333	93.333
Success_C	0	0	0	18.182	9.091	27.273	36.364	9.091	18.182
Success_D	50	100	0	0	0	0	16.667	0	50
TOTAL	43.333	45.556	36.667	53.333	65.556	70	78.889	78.889	82.222

- Euclidean, average

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	100	0	3.571
Success_B	95.556	15.556	24.444	55.556	68.889	84.444	86.667	93.333	93.333
Success_C	0	0	9.091	45.455	27.273	36.364	36.364	0	0
Success_D	33.333	100	0	0	0	0	0	0	0
TOTAL	50	45.556	44.444	64.444	68.889	77.778	78.889	46.667	47.778

Success rates - aggl. hier. clustering - 3x3

- Manhattan, Complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	0	100	100	100	100	0	100	96.429
Success_B	82.222	73.333	24.444	57.778	68.889	84.444	86.667	93.333	93.333
Success_C	0	0	54.545	45.455	36.364	36.364	0	36.364	9.091
Success_D	0	100	0	0	0	0	0	0	0
TOTAL	41.111	43.333	50	65.556	70	77.778	43.333	82.222	77.778

- Cosine, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	46.429	85.714	85.714	0	67.857	64.286	25	10.714	46.429
Success_B	75.556	55.556	60	86.667	31.111	80	91.111	95.556	95.556
Success_C	18.182	9.091	18.182	45.455	45.455	9.091	18.182	9.091	0
Success_D	0	0	0	0	0	0	0	0	0
TOTAL	54.444	55.556	58.889	48.889	42.222	61.111	55.556	52.222	62.222

- L1, complete

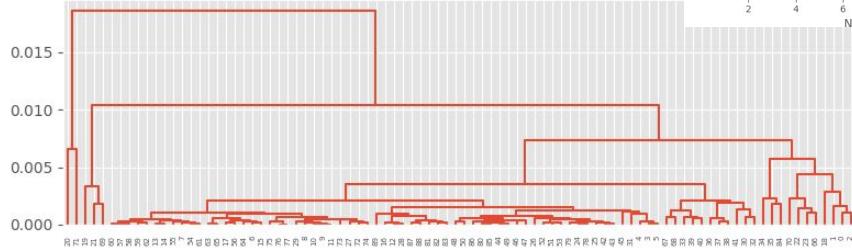
	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	0	100	100	100	100	0	100	96.429
Success_B	82.222	73.333	24.444	57.778	68.889	84.444	86.667	93.333	93.333
Success_C	0	0	54.545	45.455	36.364	36.364	0	36.364	9.091
Success_D	0	100	0	0	0	0	0	0	0
TOTAL	41.111	43.333	50	65.556	70	77.778	43.333	82.222	77.778

- L2, complete

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	100	100	100	100	100	0	96.429
Success_B	95.556	15.556	24.444	57.778	13.333	73.333	86.667	93.333	93.333
Success_C	0	0	54.545	18.182	18.182	27.273	36.364	0	9.091
Success_D	0	0	0	0	0	0	0	0	0
TOTAL	47.778	38.889	50	62.222	40	71.111	78.889	46.667	77.778

2x2

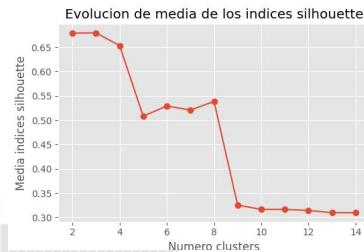
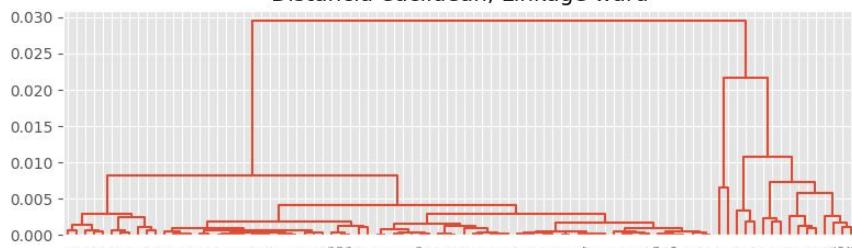
Distancia euclidean, Linkage complete



Distancia euclidean, Linkage average



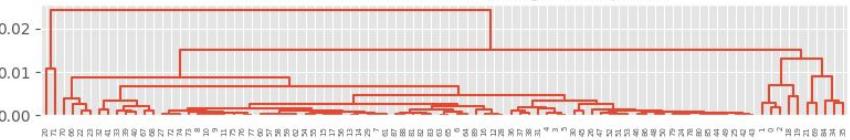
Distancia euclidean, Linkage ward



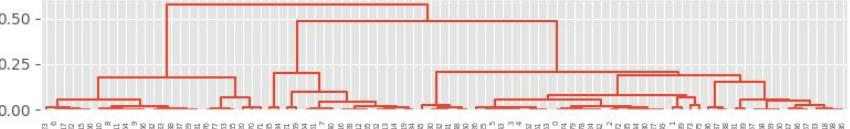
Distancia euclidean, Linkage complete



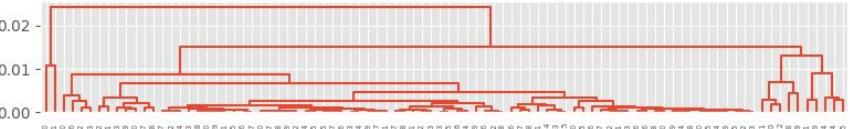
Distancia manhattan, Linkage complete



Distancia cosine, Linkage complete



Distancia l1, Linkage complete



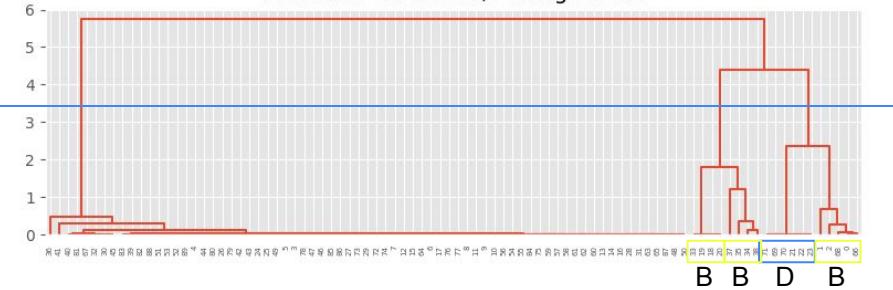
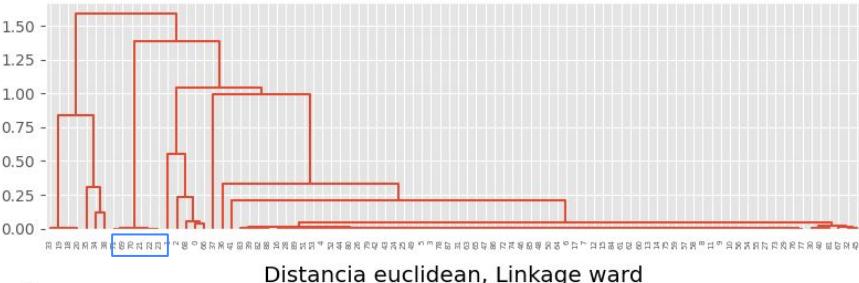
Distancia l2, Linkage complete



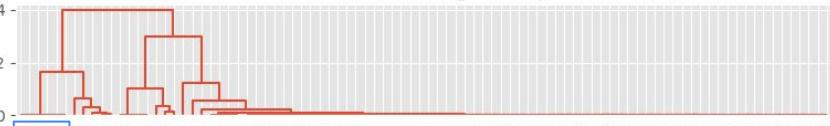
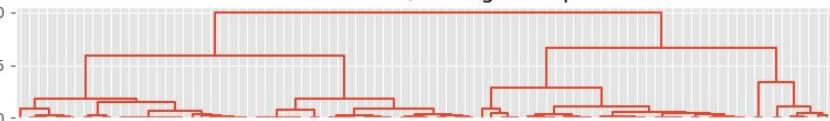
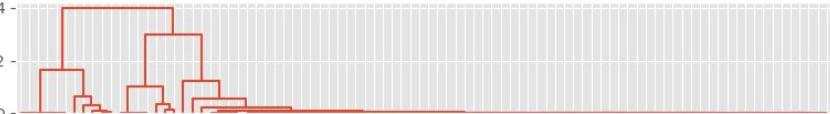
3x3



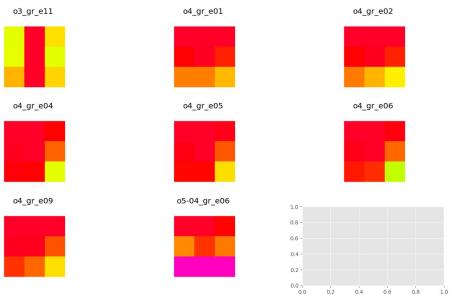
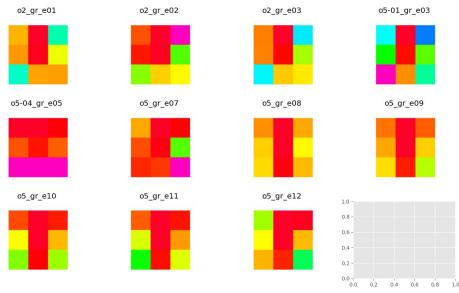
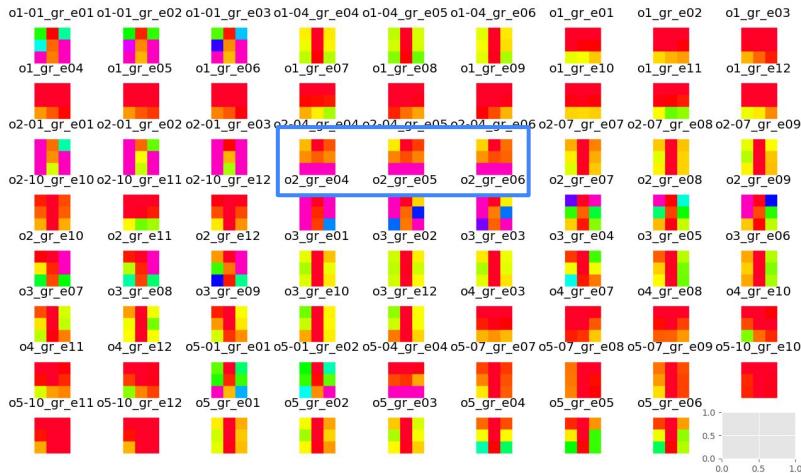
Distancia euclidean, Linkage complete



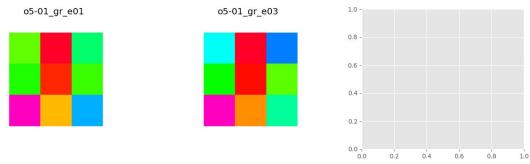
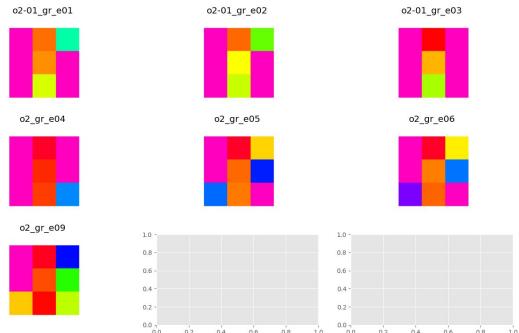
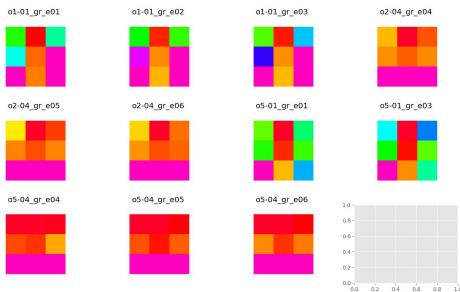
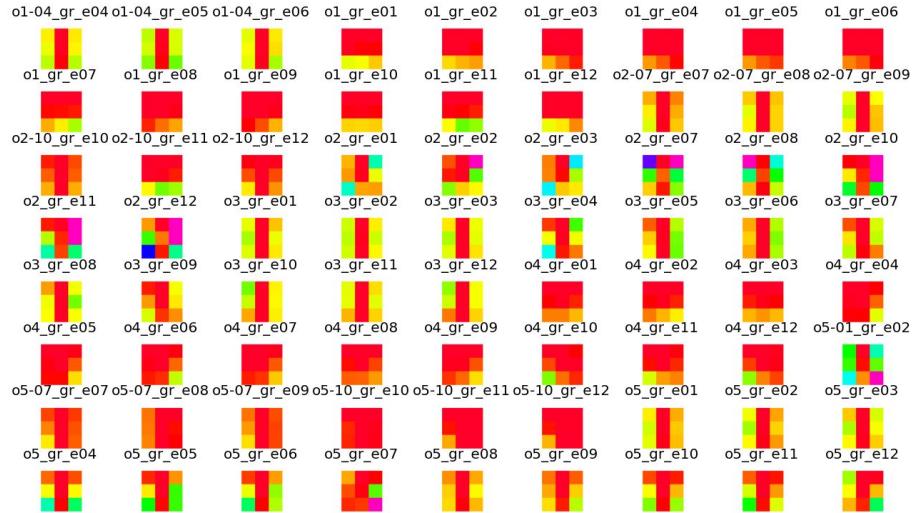
Distancia euclidean, Linkage complete



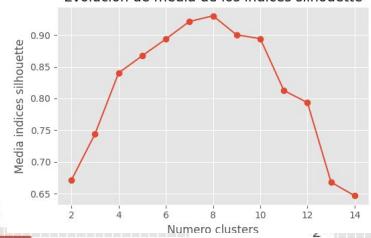
3x3 - Euclidean ward, 3 clases



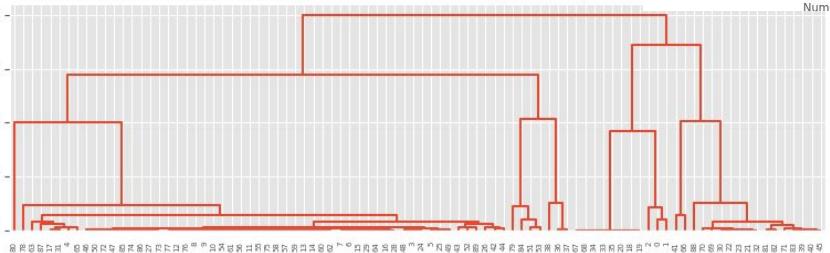
3x3 - Euclidean complete, 4 clases



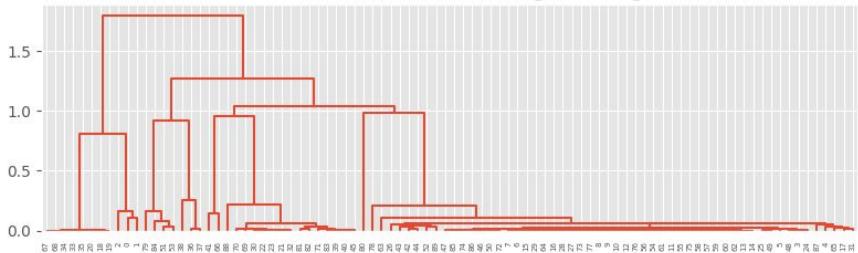
4x4



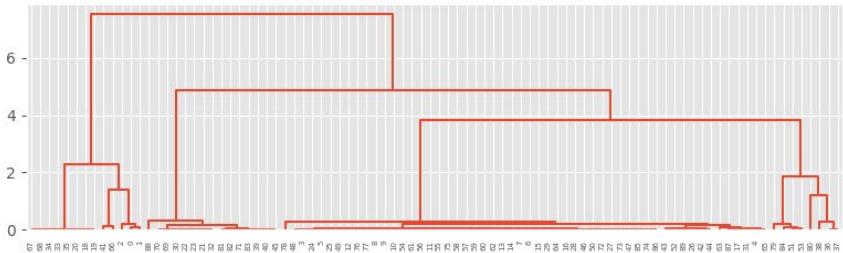
Distancia euclidean, Linkage complete



Distancia euclidean, Linkage average



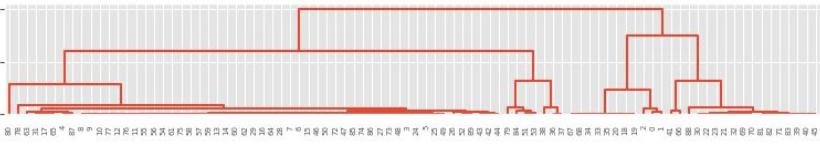
Distancia euclidean, Linkage ward



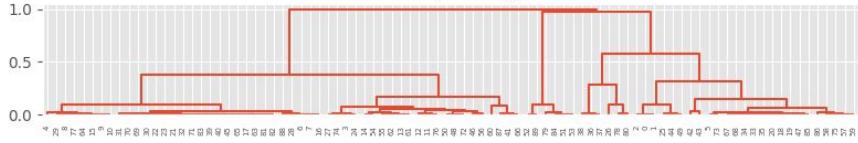
Distancia euclidean, Linkage complete



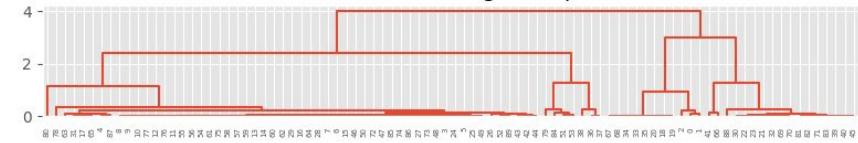
Distancia manhattan, Linkage complete



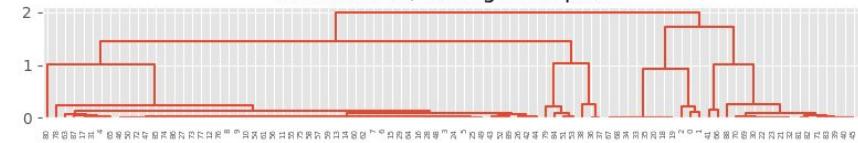
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

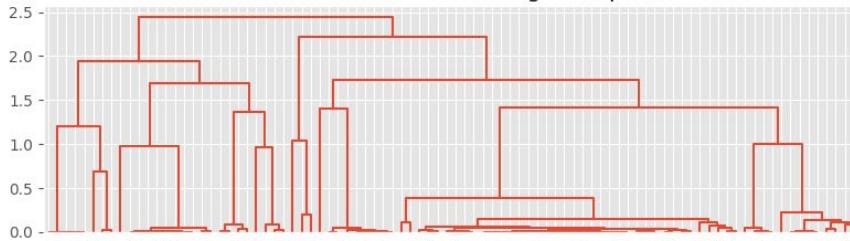


Distancia l2, Linkage complete

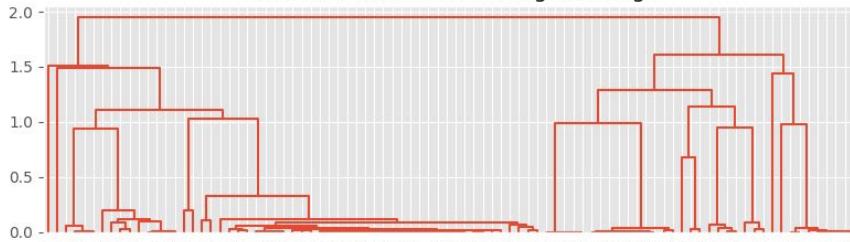


5x5

Distancia euclidean, Linkage complete



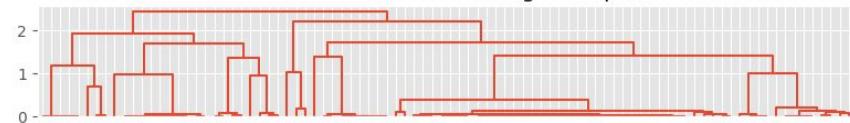
Distancia euclidean, Linkage average



Distancia euclidean, Linkage ward



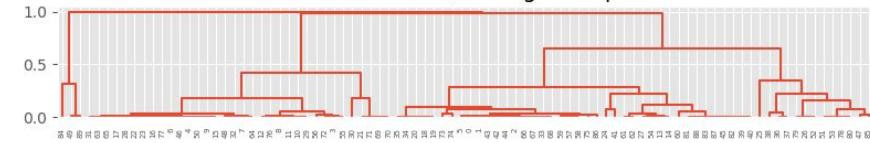
Distancia euclidean, Linkage complete



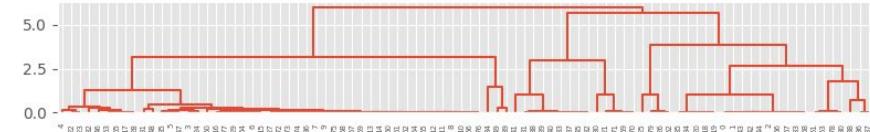
Distancia manhattan, Linkage complete



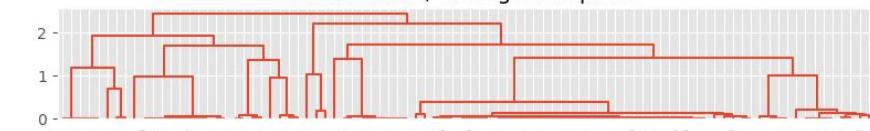
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

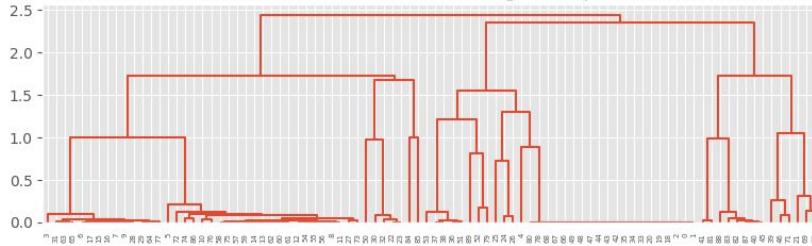


Distancia l2, Linkage complete

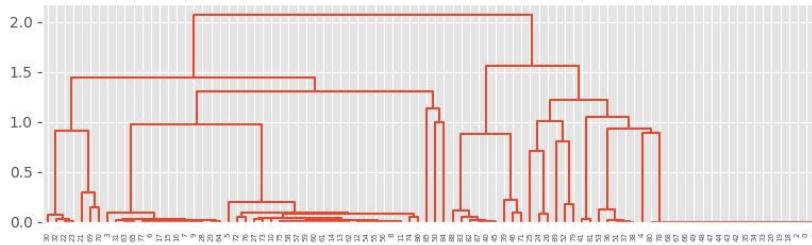


6x6

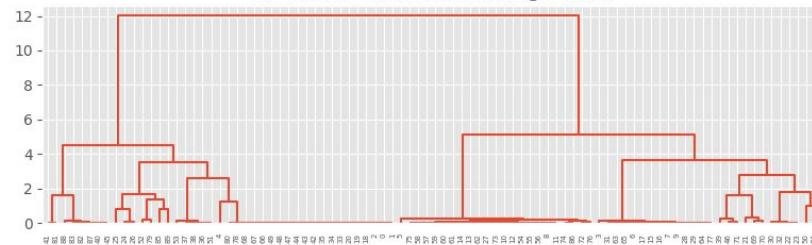
Distancia euclidean, Linkage complete



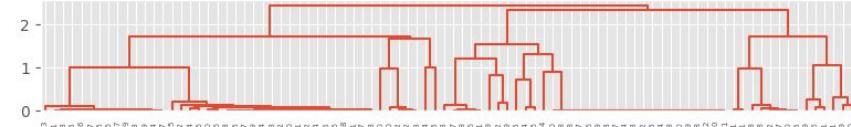
Distancia euclidean, Linkage average



Distancia euclidean, Linkage ward



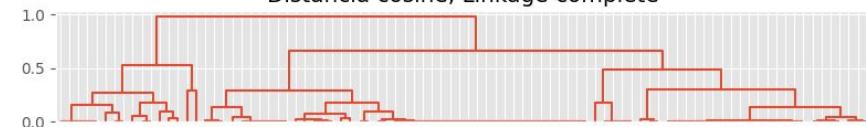
Distancia euclidean, Linkage complete



Distancia manhattan, Linkage complete



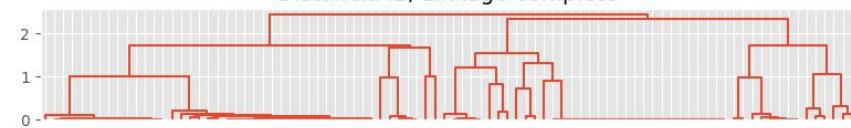
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

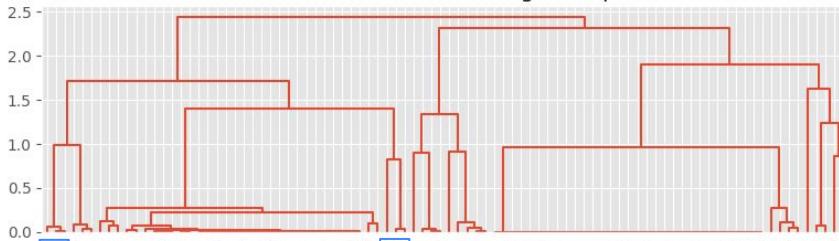


Distancia l2, Linkage complete

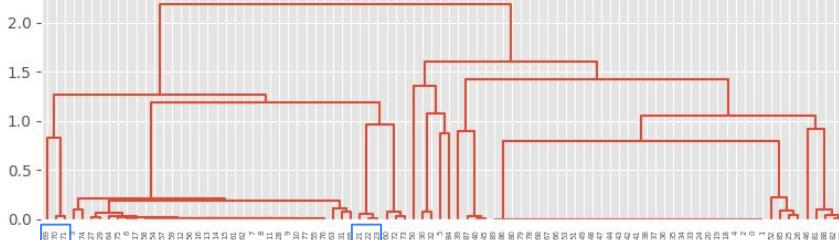


7x7

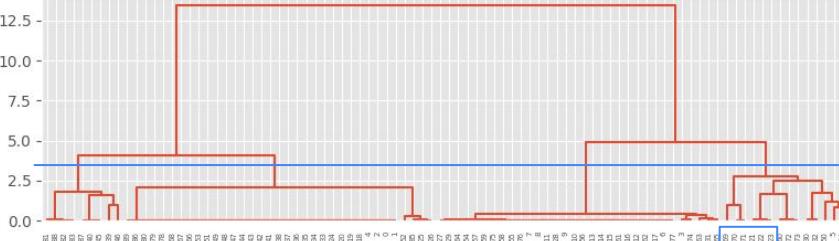
Distancia euclidean, Linkage complete



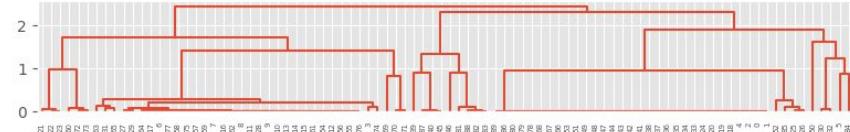
Distancia euclidean, Linkage average



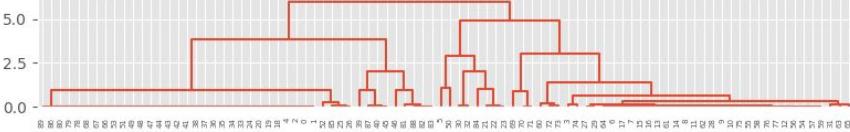
Distancia euclidean, Linkage ward



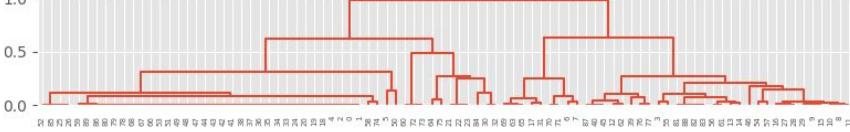
Distancia euclidean, Linkage complete



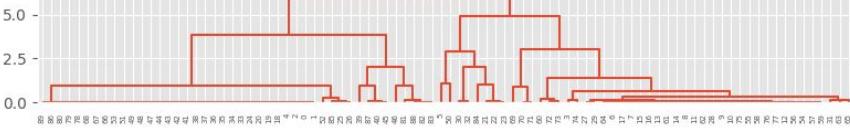
Distancia manhattan, Linkage complete



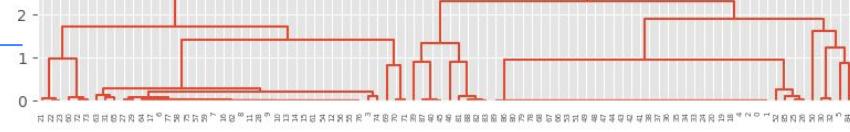
Distancia cosine, Linkage complete



Distancia l1, Linkage complete

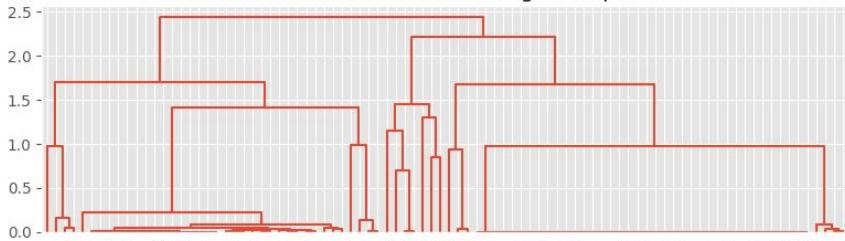


Distancia l2, Linkage complete

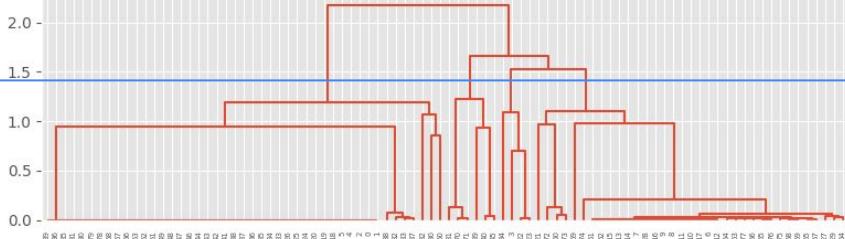


8x8

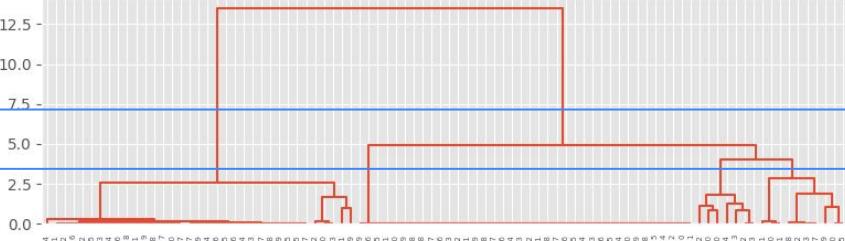
Distancia euclidean, Linkage complete



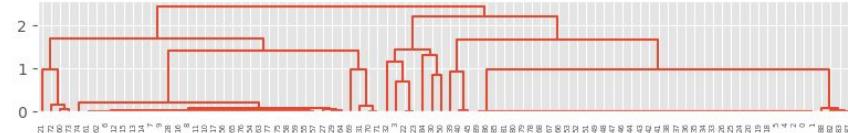
Distancia euclidean, Linkage average



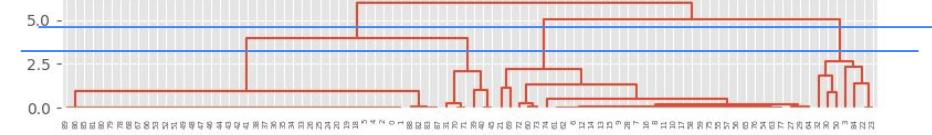
Distancia euclidean, Linkage ward



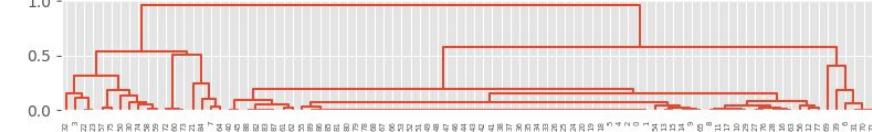
Distancia euclidean, Linkage complete



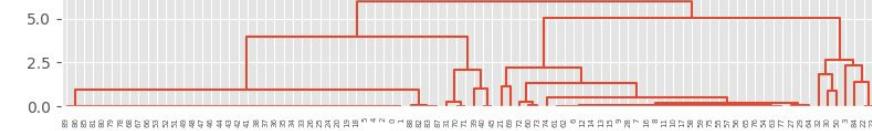
Distancia manhattan, Linkage complete



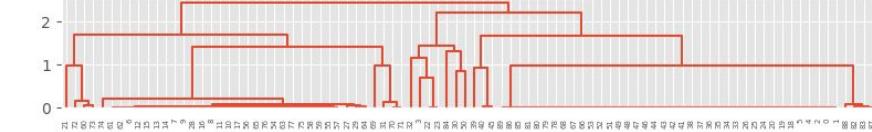
Distancia cosine, Linkage complete



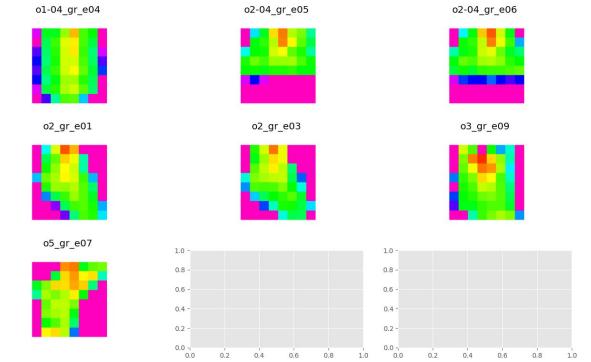
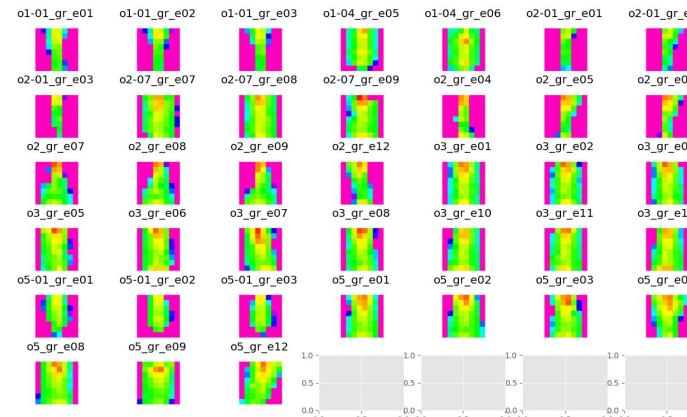
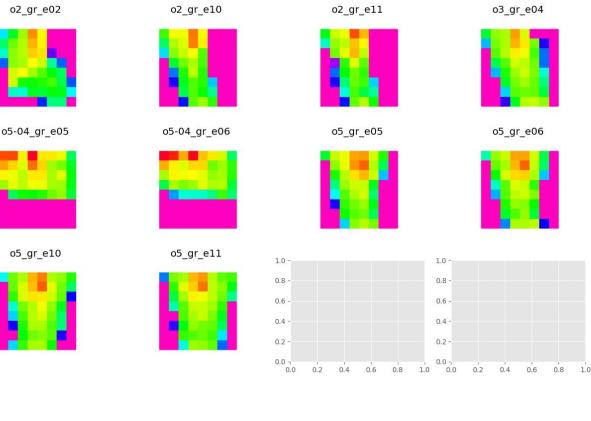
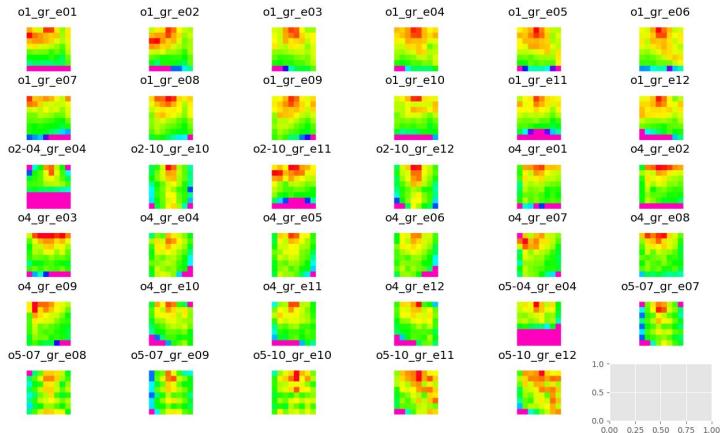
Distancia l1, Linkage complete



Distancia l2, Linkage complete



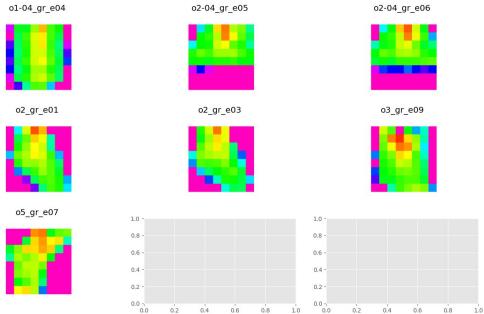
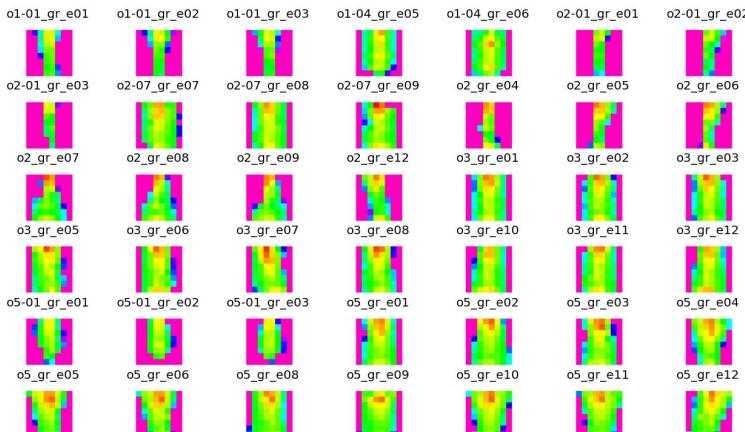
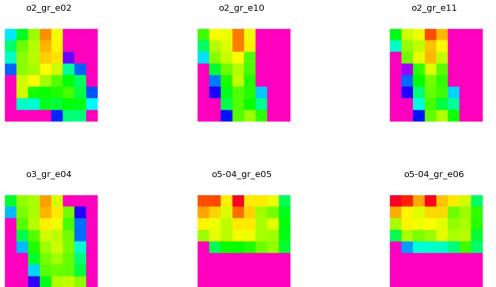
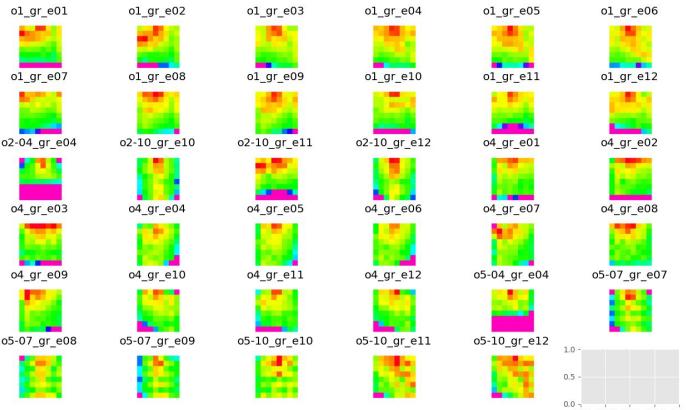
8x8 - Euclidean ward, 4 clases



8x8 - manhattan complete, 3 clases

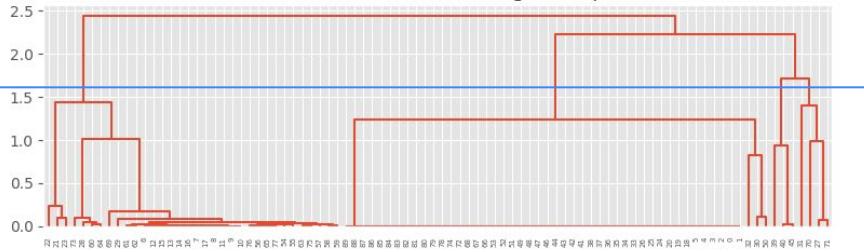


8x8 - manhattan complete, 4 clases

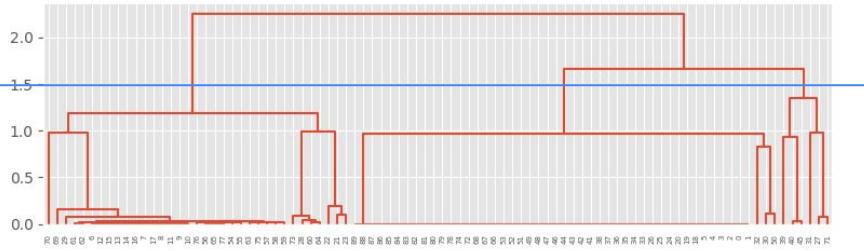


9x9

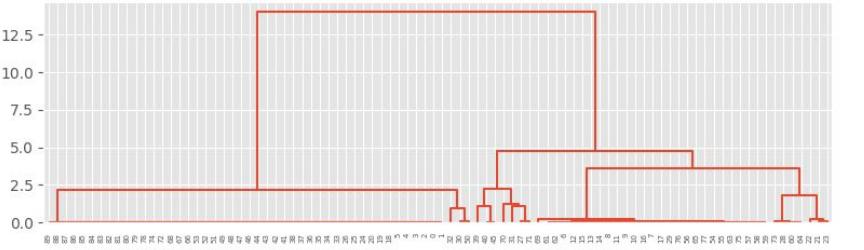
Distancia euclidean, Linkage complete



Distancia euclidean, Linkage average



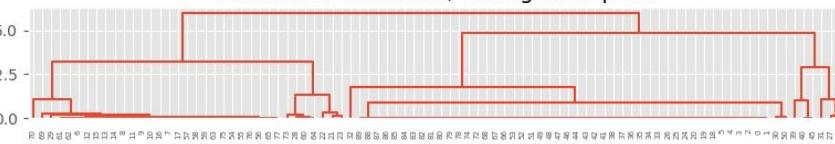
Distancia euclidean, Linkage ward



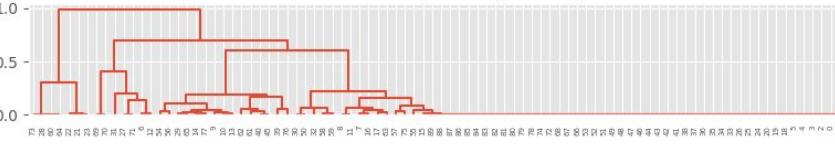
Distancia euclidean, Linkage complete



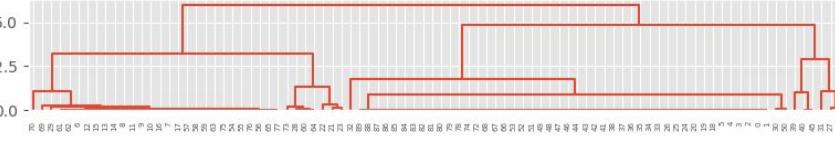
Distancia manhattan, Linkage complete



Distancia cosine, Linkage complete



Distancia l1, Linkage complete

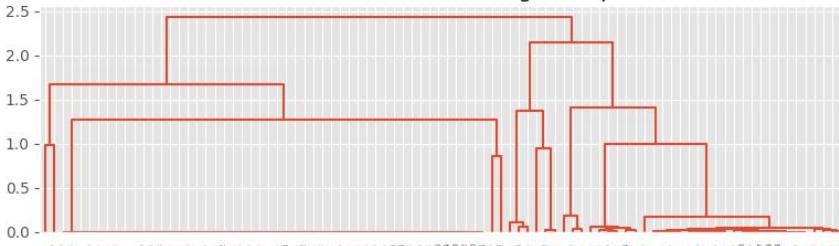


Distancia l2, Linkage complete

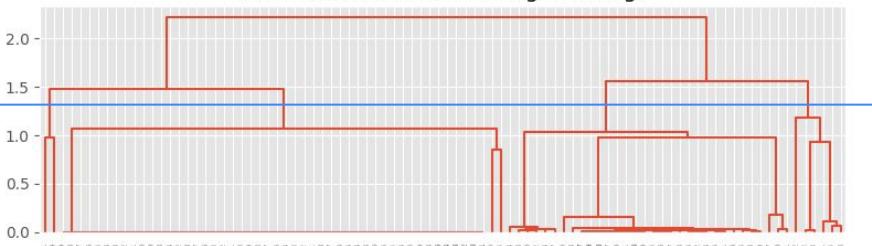


10x10

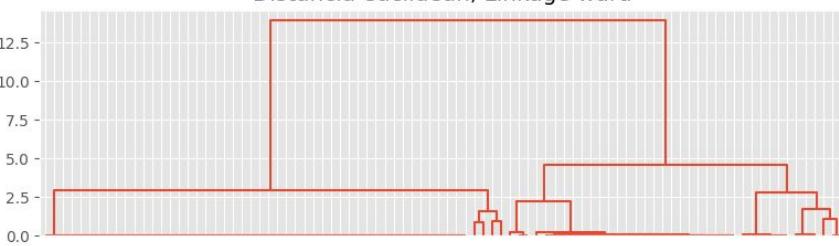
Distancia euclidean, Linkage complete



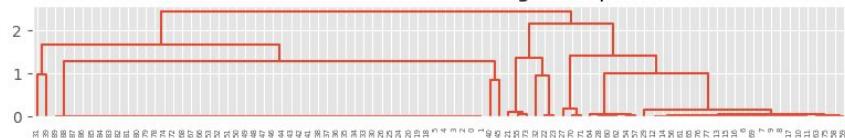
Distancia euclidean, Linkage average



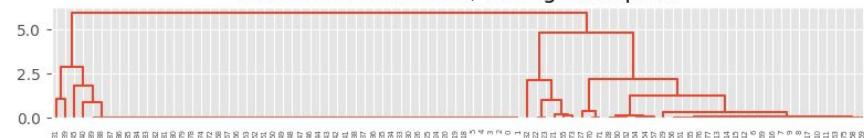
Distancia euclidean, Linkage ward



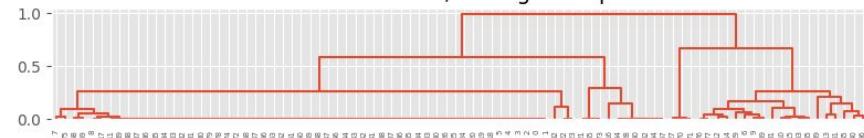
Distancia euclidean, Linkage complete



Distancia manhattan, Linkage complete



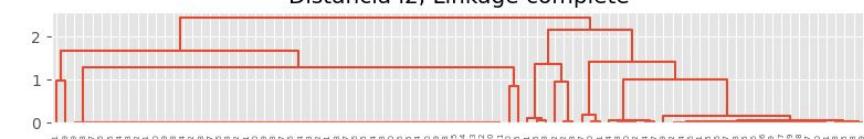
Distancia cosine, Linkage complete



Distancia l1, Linkage complete



Distancia l2, Linkage complete



Kmeans

Combining class label assignment and selecting the one that maximises total success

- n_clusters=4, init='random', n_init=20, max_iter=500, tol=1e-8, random_state=0

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	0	100	0	92.857	100	100	100	100	96.429
Success_B	73.333	15.556	57.778	40	0	57.778	0	62.222	55.556
Success_C	45.455	36.364	27.273	0	27.273	27.273	63.636	0	27.273
Success_D	33.333	100	100	100	0	100	0	0	100
TOTAL	44.444	50	38.889	55.556	34.444	70	38.889	62.222	67.778

- n_clusters=4, init=k-means++, n_init=20, max_iter=500, tol=1e-8, random_state=0

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	92.857	100	100	100	100	0
Success_B	20	20	26.667	44.444	60	26.667	26.667	26.667	62.222
Success_C	0	36.364	0	0	0	54.545	72.727	72.727	0
Success_D	33.333	100	100	100	0	100	100	100	100
TOTAL	43.333	52.222	51.111	57.778	61.111	57.778	60	60	37.778

- n_clusters=4, init=[init_central], n_init=20, max_iter=500, tol=1e-8, random_state=None -> samples with LESS intra dist to centroid of class

	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Success_A	100	100	100	96.429	100	100	100	100	100
Success_B	13.333	11.111	33.333	37.778	62.222	82.222	88.889	55.556	55.556
Success_C	36.364	36.364	36.364	45.455	63.636	27.273	63.636	27.273	27.273
Success_D	50	100	100	100	100	100	100	100	100
TOTAL	45.556	47.778	58.889	61.111	76.667	82.222	90	68.889	68.889

Reordenar samples segun clase