

Análisis de Datos 1

Sumario

Aprendizaje automático supervisado (clasificación, regresión)

Análisis discriminante

Lineal (`MASS::lda`)

Cuadrático (`MASS::qda`)

Árboles de decisión (`rpart::rpart`)

Redes neuronales (`nnet::nnet`)

Aprendizaje no supervisado (minería de datos, análisis exploratorio)

Análisis de conglomerados (clúster)

Métodos jerárquicos (`hclust`)

Particionado óptimo (`kmeans`)

Métodos de visualización

Gráficas de dispersión por parejas (`pairs`)

Componentes principales (`princomp`, `prcomp`; `MASS::corresp`)

Métodos de distancia

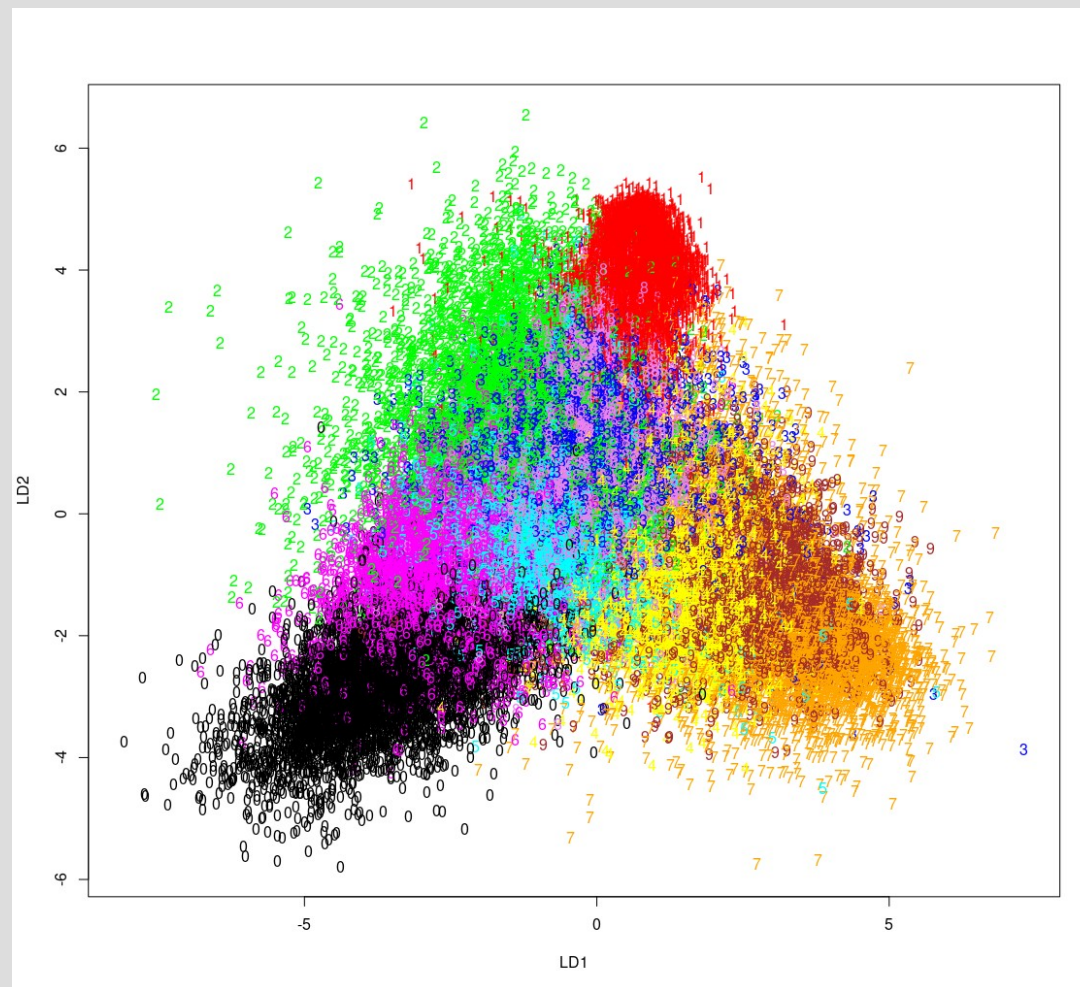
Escalado multidimensional (`cmdscale`)

Mapeo no lineal (`MASS::sammon`)

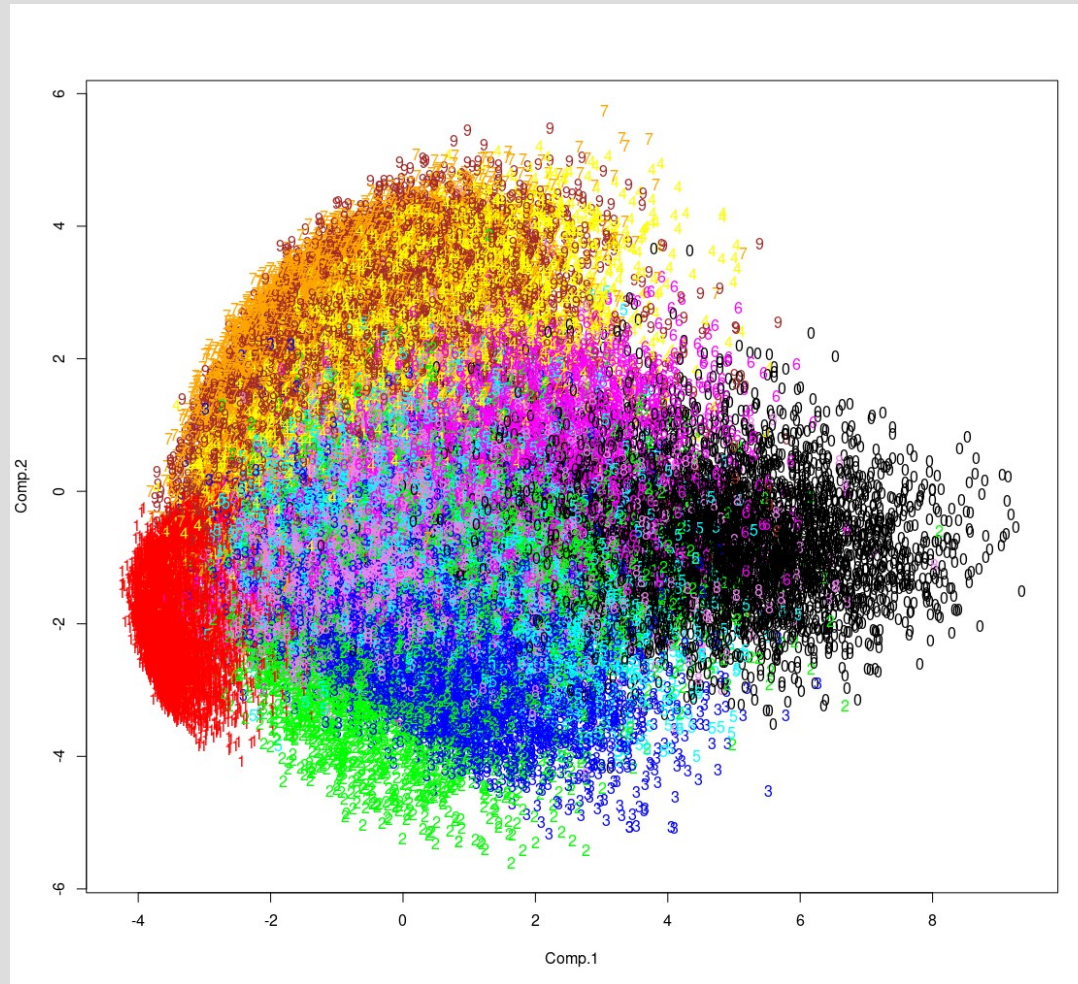
Mapas auto-organizados (`class::batchSOM`, `SOMbrero::trainSOM`)

t-SNE (`Rtsne::Rtsne`)

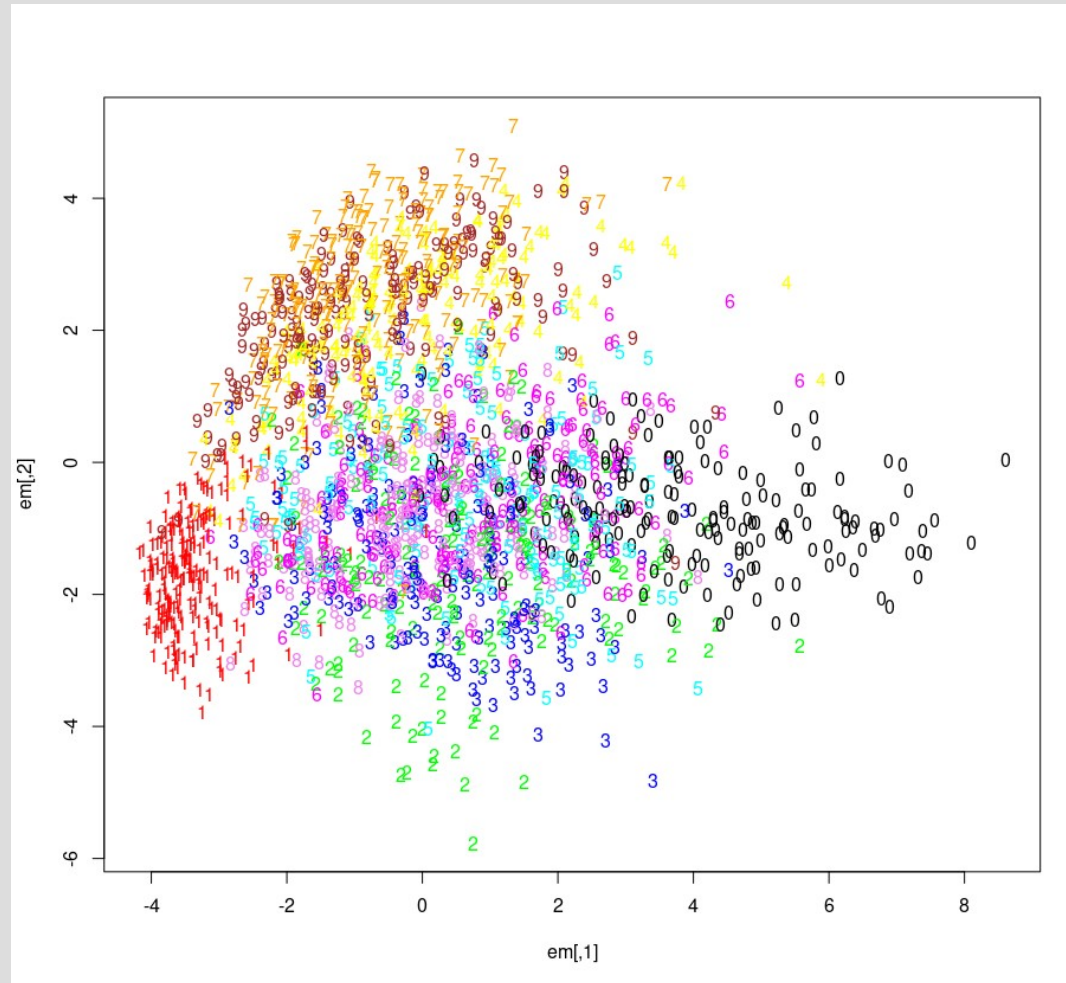
Discriminante lineal



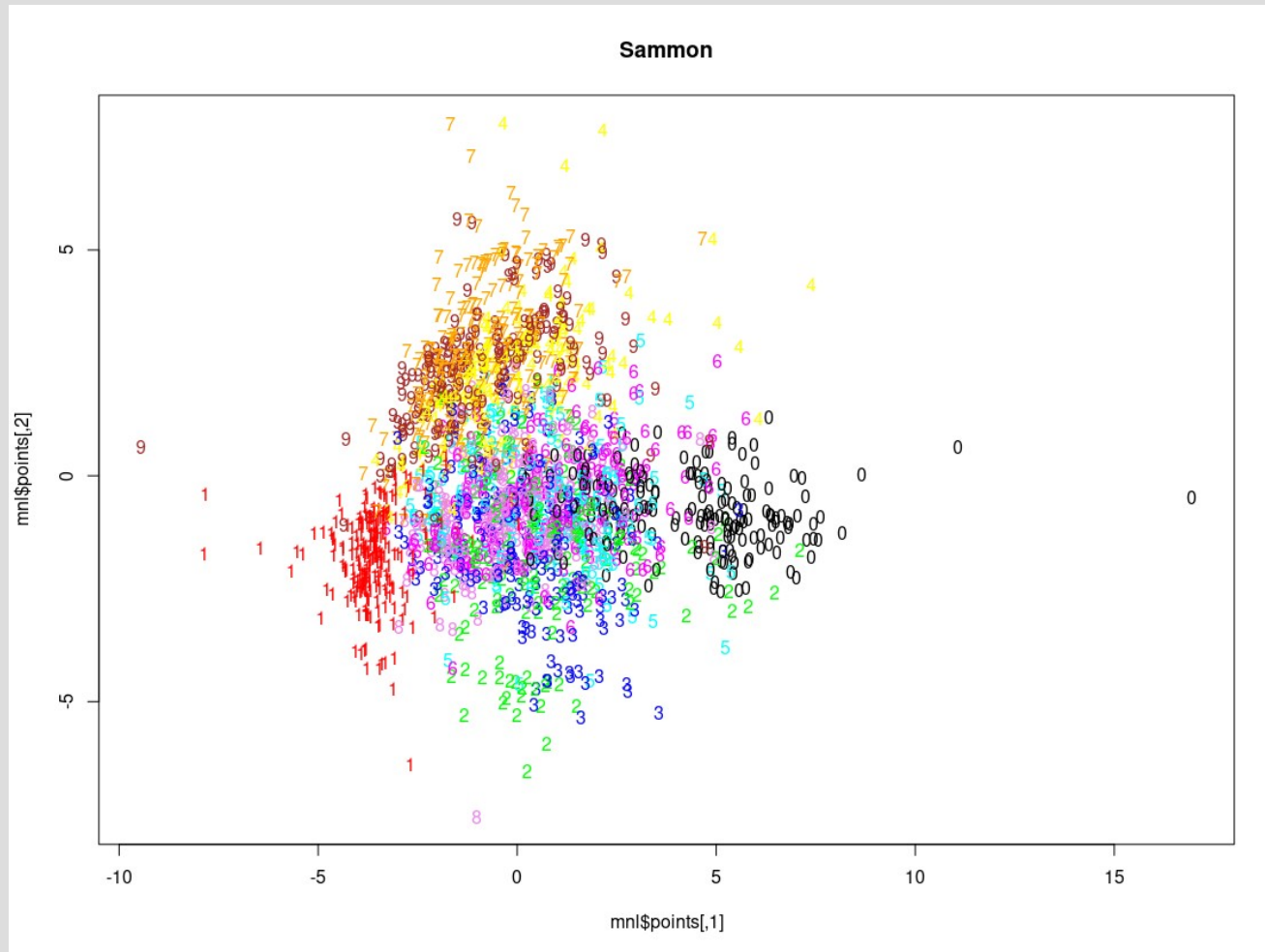
Componentes principales



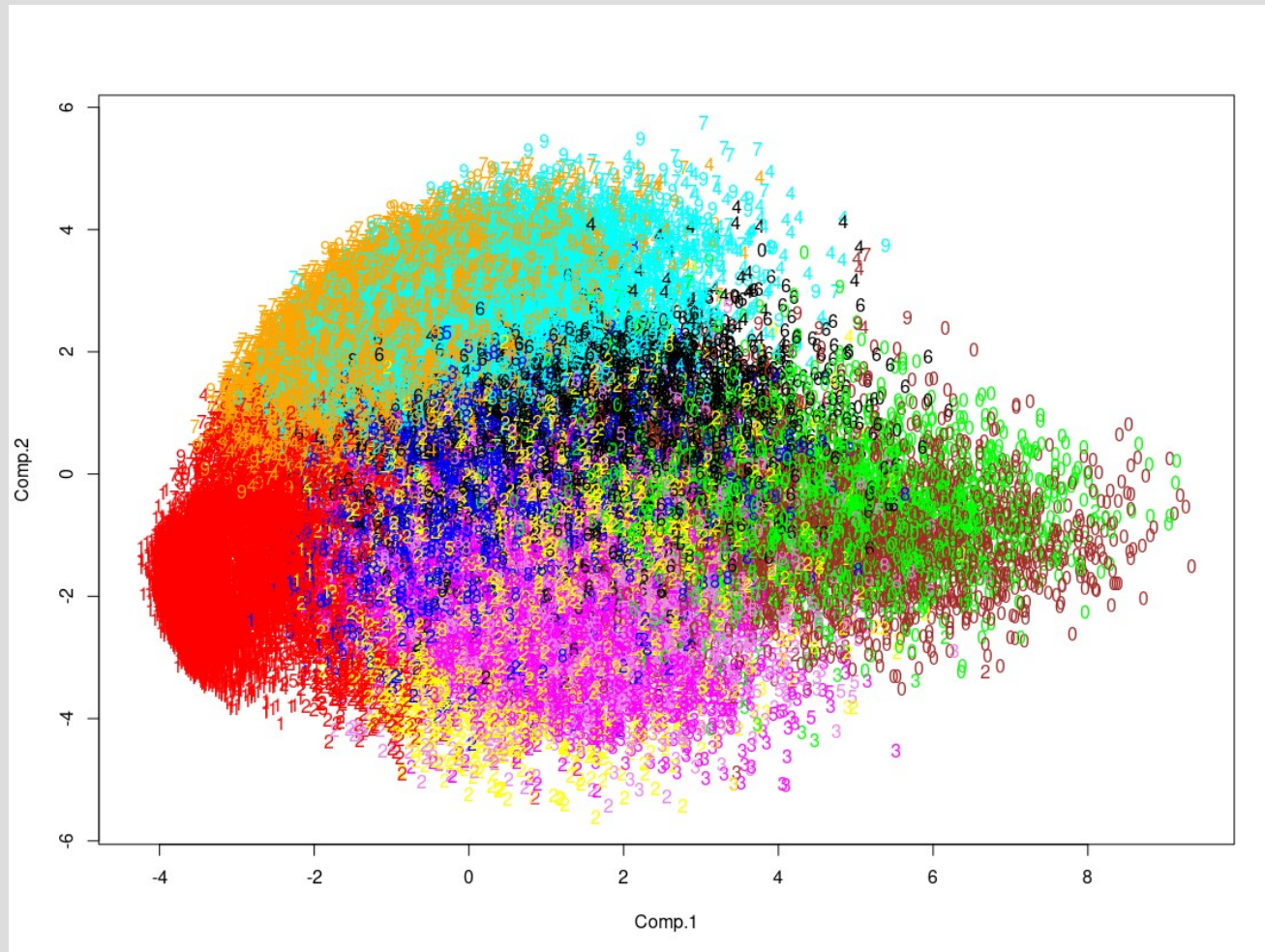
Escalado multidimensional



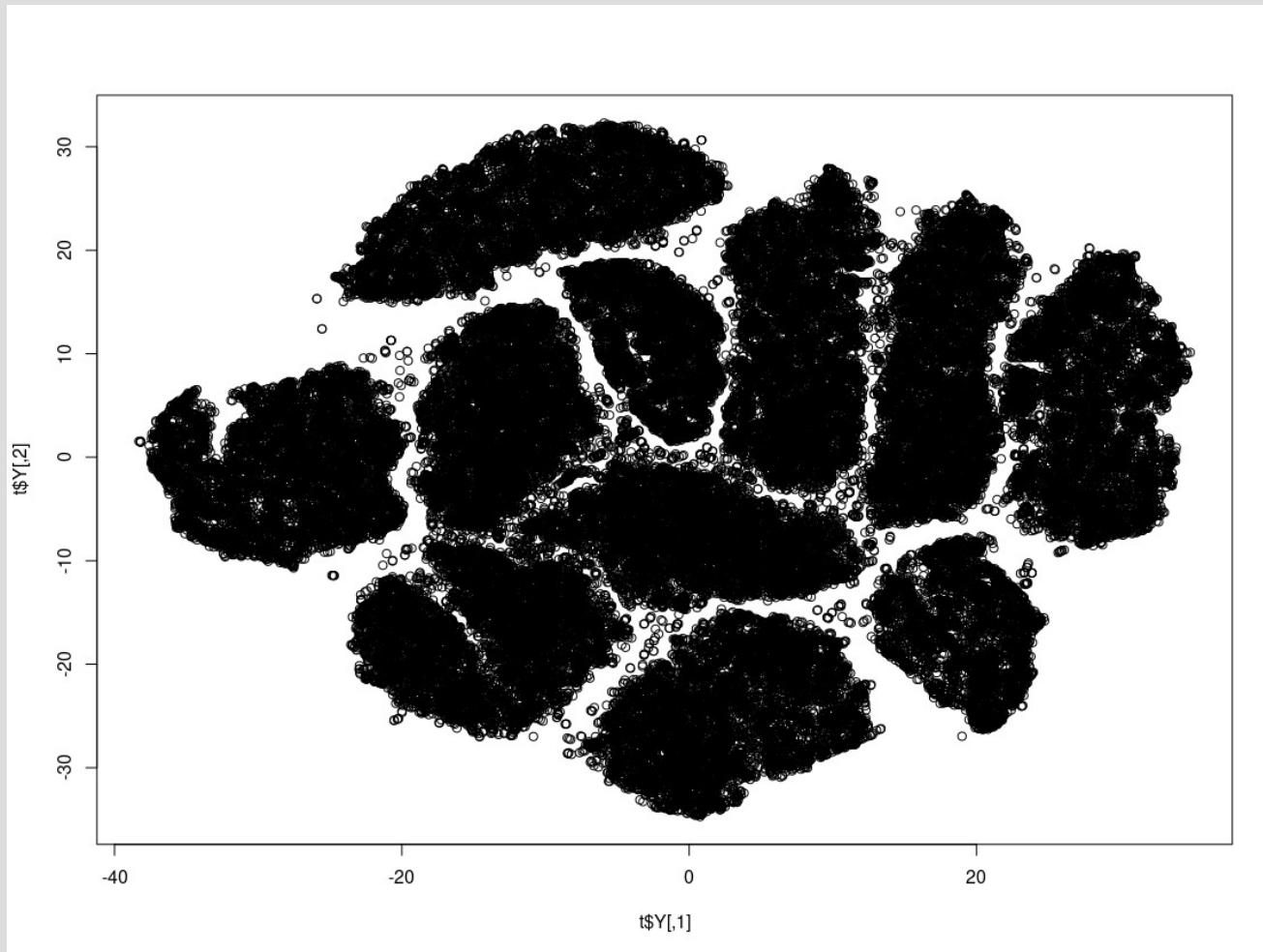
Mapeo no lineal



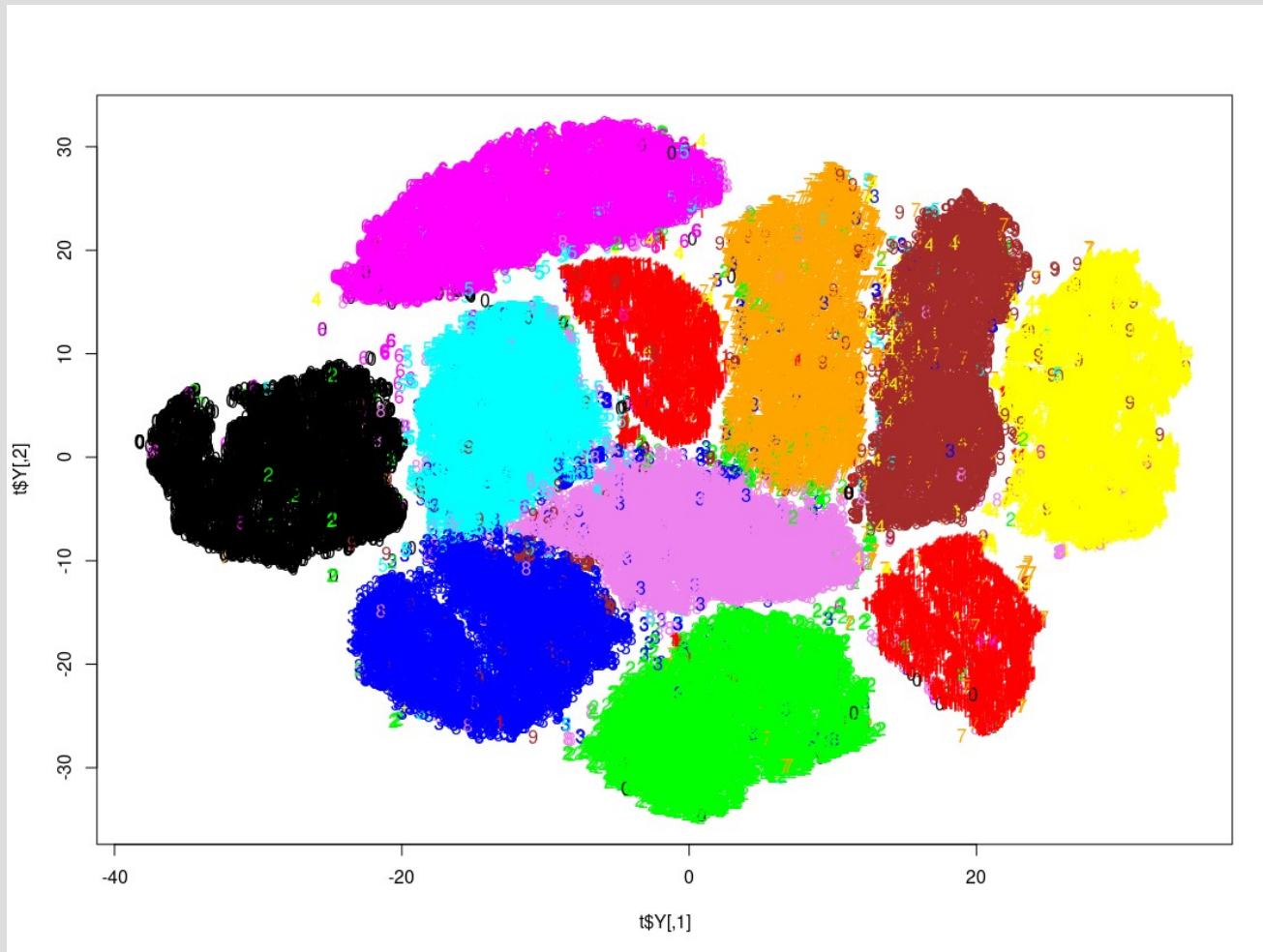
k-Medias



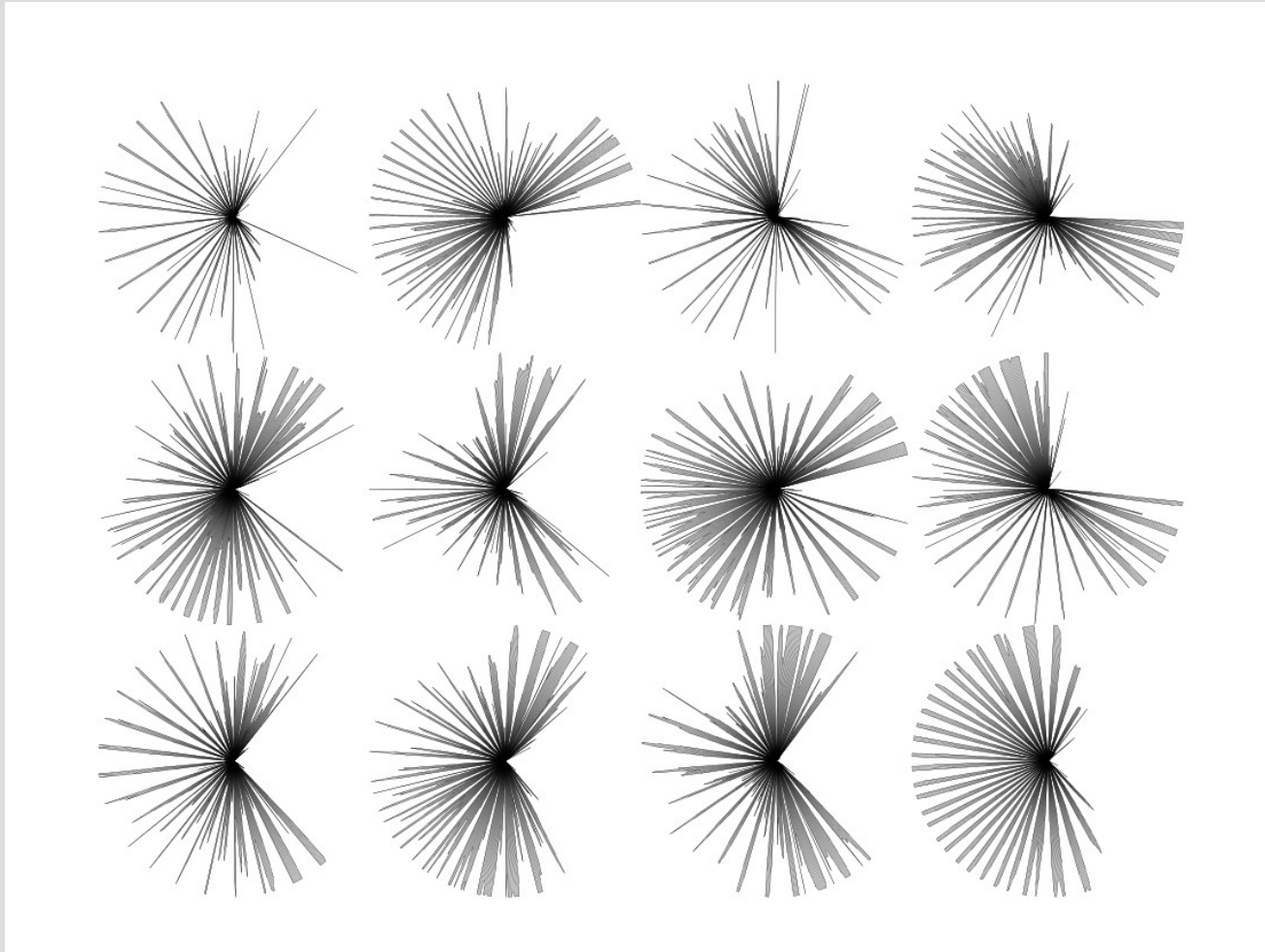
t-SNE



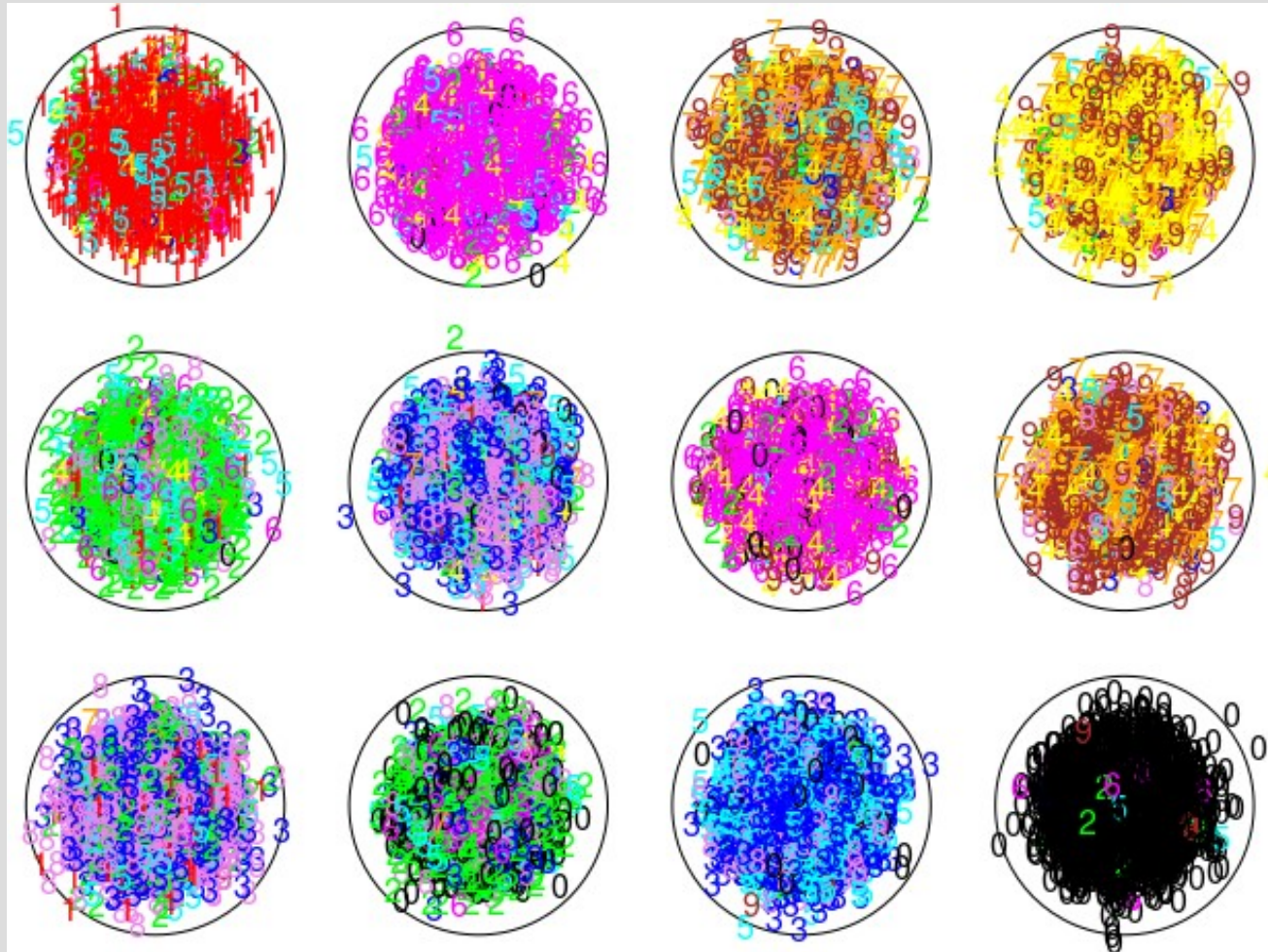
t-SNE



Mapas autoorganizados



Mapas autoorganizados



Mapas autoorganizados

real \ predicho

	1	2	3	4	5	6	7	8	9	10	11	12
0	29	1113	452	4680	63	159	229	19	2	126	10	21
1	413	0	0	0	344	149	1	1	6924	12	33	0
2	765	1656	50	31	3098	429	227	20	278	252	115	69
3	1666	441	2577	10	142	1890	10	114	99	51	131	10
4	12	16	1	1	90	92	735	728	94	445	1122	3488
5	411	321	1998	23	640	1007	33	299	351	244	743	243
6	17	304	30	40	409	57	2114	0	143	3750	4	8
7	49	12	1	17	103	86	47	2175	217	5	3199	1382
8	2606	281	705	10	471	1654	50	348	212	66	270	152
9	29	18	54	20	29	205	424	2448	66	21	2260	1384

Escalado multidimensional

$$\frac{\sum_{i \neq j} (d_{ij} - D_{ij})^2}{\sum_{i \neq j} d_{ij}^2}$$

Mapeo no lineal de Sammon

$$\frac{1}{\sum_{i \neq j} d_{ij}} \sum_{i \neq j} \frac{(d_{ij} - D_{ij})^2}{d_{ij}}$$

t-SNE

Calcular probabilidades

$$p_{j|i} \propto e^{-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}}$$

Definir distancias en el espacio de origen

$$d_{ij} \propto p_{j|i} + p_{i|j}$$

Definir distancias en el espacio de destino

$$D_{ij} \propto \frac{1}{1 + \|y_i - y_j\|^2}$$

Buscar y_i que hagan mínima la divergencia KL

$$KL = \sum_{i \neq j} d_{ij} \log \frac{d_{ij}}{D_{ij}}$$