

Reconocimiento estadístico de patrones - Proyecto 01
Índice de marginación en México en el periodo 2015 - 2020
Giovanni Gamaliel López Padilla

Resumen

El índice de marginación busca representar las necesidades básicas insatisfechas utilizando una baja dimensionalidad en los datos. El Consejo Nacional de Población (CONAPO) creó este índice en el año 1990 empleando indicadores para recabar información acerca de la situación socioeconómica de los municipios y entidades. A partir de estos indicadores socioeconómicos se emplea el algoritmo de análisis de componentes principales (PCA) para obtener una representación en una sola dimensión y así clasificar el estado del municipio o entidad. En este trabajo se exploran alternativas dentro de los algoritmos de reducción de dimensionalidad y de clasificación. Resultando así que para casos unidimensionales, el algoritmo PCA con kernel lineal o sigmoide son buenas alternativas para realizar la representación y que los algoritmos de clasificación no son una opción para la tarea debido a la baja presición de resultados.

Índice

1. Introducción	2
2. Métodos y materiales	4
2.1. Algoritmos de reducción de dimensionalidad	4
2.1.1. ISOMAP	4
2.1.2. T-SNE	4
2.1.3. Análisis de componentes principales	5
2.2. Algoritmos de clasificación	6
2.2.1. K-means	6
2.2.2. Hierarchical Cluster	6
2.2.3. Self-Organizing-Map	6
2.3. Base de datos de CONAPO	7
3. Resultados	8
3.1. Algoritmos de reducción de dimensionalidad	8
3.1.1. ISOMAP	8
3.1.2. T-SNE	10
3.1.3. Análisis de componentes principales	11
3.2. Algoritmos de clasificación	12
3.2.1. K-means	12
3.2.2. Hierarchical Cluster	13
3.2.3. Self-Organizing-Map	13

4. Conclusiones	14
4.1. Algoritmos de reducción de dimensionalidad	14
4.2. Algoritmos de clasificación	15
5. Código	15
6. Referencias	15

1. Introducción

Uno de los problemas que se presentan en las sociedades humanas es el difícil acceso a los servicios públicos como salud, educación o justicia. La deficiencia de los servicios públicos propician condiciones de vulnerabilidad e inseguridad. Una consecuencia de estas condiciones de vida es la disminución de la calidad de vida de los habitantes. Con este problema surgen instituciones nacionales e internacionales para estudiar el caso, recabar información y generar indicadores que ayuden a identificar los casos en una región extensa. El consejo Nacional de Población (CONAPO) calcula el Índice de Marginación (IM) basado en el fundamento en el enfoque de necesidades básicas insatisfechas, utilizando el análisis de componentes principales¹. De una manera semejante, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) estima el Índice de Rezago Social (IRS) por medio de una suma ponderada de porcentajes². El IM se utiliza para localizar las zonas de atención prioritaria. Estas zonas tienen acceso a fondos y programas gubernamentales para el combate de la marginación y el desarrollo de la población³⁻¹¹. Los programas gubernamentales que usan al IM para determinar las localidades que serán beneficiarias son:

- Estrategia 100x100¹².
- Estrategia del Gobierno Federal para la Dotación de Piso Firme¹³.
- Programa de Apoyo Alimentario en Zonas de Atención Prioritaria¹⁴.
- Programa de Apoyo a Zonas de Atención Prioritaria¹⁵.

El IM utiliza como indicadores¹⁶ los siguientes porcentajes:

- Población analfabeta de 15 años o más.
- Población de 15 años o más sin educación básica.
- Viviendas particulares sin agua entubada.
- Viviendas particulares sin drenaje ni excusado.
- Viviendas particulares sin energía eléctrica.
- Viviendas particulares con piso de tierra.
- Viviendas particulares con hacinamiento
- Porcentaje de población que vive en localidades menores a 5000 habitantes.
- Población ocupada con ingresos de hasta 2 salarios mínimos.

Para estimar los índices de marginación se utiliza el método de distancia P_2^{17} (DP_2). El método de distancia P_2 fue desarrollado para hacer comparaciones temporales y espaciales. Ha sido aplicado en investigaciones sobre la calidad de vida en la Unión Europea y España¹⁸⁻²¹. La base del método es tener una base de referencia. Los valores de referencia pueden conformarse por los mínimos o máximos para todas los indicadores o para cada indicador. Estos valores no tienen que ser necesariamente reales. Si se toman los valores mínimos, entonces el indicador de DP_2 tiene valores altos cuando existen mejores condiciones socioeconómicas. La base de referencia a nivel municipal para el periodo 2010-2020 se muestra en la tabla 1¹⁶. El DP_2 evita la duplicación de información e imparcialidad en el esquema de ponderación. Además de esto contiene las siguientes propiedades matemáticas: existencia, determinación monotonía, unicidad, invarianza frente a la base de referencia, homogeneidad, transitividad, exhaustiva, aditividad y neutralidad^{18;22}. Debido a las propiedades presentes en el indicador DP_2 dan un beneficio al análisis de componentes principales¹⁸.

Indicadores socioeconomics	Base de referencia
Porcentaje de población analfabeta de 15 años o más.	-66.74
Porcentaje de población de 15 años o más sin educación básica.	-94.79
Porcentaje de viviendas particulares sin agua entubada.	-89.90
Porcentaje de viviendas particulares sin drenaje ni excusado.	-69.45
Porcentaje de viviendas particulares sin energía eléctrica.	-99.74
Porcentaje de viviendas particulares con piso de tierra.	-79.71
Porcentaje de viviendas particulares con hacinamiento	-83.24
Porcentaje de porcentaje de población que vive en localidades menores a 5000 habitantes.	-100.00
Porcentaje de población ocupada con ingresos de hasta 2 salarios mínimos.	-100.00

Tabla 1: Base de referencia a nivel municipal en el periodo 2010-2020.

El indicador DP_2 sintético se define en la ecuación 1.

$$DP_2 = \sum_i \frac{d_{ij}}{\sigma_j} (1 - R_{j,j-1,\dots,1}^2) \quad R_1^2 = 0 \quad (1)$$

donde d_{ij} es la distancia de la j -ésima variable del municipio con respecto a la base de referencia, σ_j es la desviación estándar de la variable j y $R_{j,j-1,\dots,1}^2$ es el coeficiente de determinación de la regresión del indicador parcial j con los demás indicadores. R_1^2 es igual a cero porque el primer indicador aporta toda la información y al no existir algún indicador previo su ponderación es la unidad. Una forma de observar el índice de marginación es normalizando (ecuación 2) sus valores para apreciar la evolución de cada municipio²³.

$$\bar{DP}_2^i = \frac{DP_2^i - \min(DP_2)}{\max(DP_2) - \min(DP_2)} \quad (2)$$

Es importante aclarar que el índice de marginación normalizado se puede tomar como una guía ya que es sensible a los valores atípicos, lo que en cierta medida hace que se amplifiquen los datos normalizados ante los casos más extremos. Pero debido a las propiedades del método y que el rango de cada uno de los índices es estrecho y lineal, se podría esperar que la normalización no genere ruido en las colas de la distribución. Cada indicador es normalizado por medio de la ecuación 3.

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (3)$$

Donde X_{ij} representa al i-ésimo valor de la variable k , μ_k y σ_k es el promedio y desviación estándar de la variable k respectivamente. Debido a la normalización de cada indicador, el

IM no puede utilizarse para realizar un seguimiento temporal. Para representar a los nuevos indicadores en uno solo se aplica el análisis de componentes principales. La CONAPO emplea el método de Dalenius Hodges²⁴ para agrupar los puntajes en cinco categorías de marginación (muy alta, alta, media, baja y muy baja). El IM permite ordenar y clasificar según su categoría de marginación en el año que se calcula.

2. Métodos y materiales

2.1. Algoritmos de reducción de dimensionalidad

La reducción de dimensionalidad es un proceso que disminuye el tamaño de variables o que describen a un conjunto de datos. Las ventajas de la reducción de dimensionalidad son las siguientes:

- Reduce el espacio de tiempo y almacenamiento requerido.
- La eliminación de multicolinealidad mejora el rendimiento del modelo de aprendizaje automático.
- Se hace más fácil de visualizar los datos cuando se reduce dos o tres dimensiones.

Los algoritmos que se implementaron en este trabajo fueron ISOMAP, K-means, Self-Organizing-Map (SOM), T-SNE y análisis de componentes principales (PCA).

2.1.1. ISOMAP

El algoritmo ISOMAP está basado en la teoría espectral, la cual trata de preservar la distancia euclídea en la reducción de dimensionalidad a la cual se quiere llegar. El método empieza creando un grafo de vecinos. En seguida se calcula la geodésica entre cada par de elementos de la base de datos, creando así una matriz de disimilaridad. A partir de la matriz de disimilaridad se emplea el escalamiento multidimensional.

2.1.2. T-SNE

El algoritmo de T-SNE consiste en crear una distribución de probabilidad que representante las similitudes entre vecinos en un espacio de gran dimensión en un espacio de menor dimensión. Para cada elemento del conjunto de datos se centra una distribución gaussiana alrededor del elemento. En seguida se obtiene la densidad bajo la distribución y normalizamos el valor, calculando así una probabilidad condicional (ecuación 4).

$$P_i = P_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/\sigma)}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2/\sigma)} \quad (4)$$

El valor de σ se define por medio de un parámetro llamado perplejidad, el cual corresponde al número de vecinos alrededor de cada punto. A una mayor probabilidad condicional, los elementos ij son más similares. A partir de las probabilidades condicionales se tratan encontrar pares de elementos tales que las distribuciones $P_{i|j}$ y $q_{i|j}$ se parecen. El valor de $q_{i|j}$ se obtiene mediante la distribución T (ecuación 5).

$$Q_i = q_{j|i} = \frac{\exp(-\|x_i - x_j\|^2)}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2)} \quad (5)$$

Por lo tanto, el algoritmo T-SNE se centra en minimizar la distancia Kullback-Leiber (ecuación 6)

$$\min d_i(P_i; Q_i) \quad d(P^1; P^2) = \sum_i P_i^1 \log \left(\frac{P_i^1}{P_i^2} \right) \quad (6)$$

2.1.3. Análisis de componentes principales

El análisis de componentes principales (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. La idea del algoritmo es asociar a cada elemento de la base de datos con un vector de menor dimensión tal que se minimiza la ecuación 7.

$$\sum_i \sum_j (d(x_i, x_j)^2 - d(x_i^* - x_j^*)^2)^2 \quad (7)$$

El problema de minimizar la ecuación 7 se puede transformar a la ecuación 8.

$$\min \left\| -\frac{1}{2} \mathbb{C}(\mathbb{D}^2 - \mathbb{D}^{*2}) \mathbb{C} \right\|_F^2 \quad (8)$$

Donde \mathbb{C} es la matriz para centrar. Definiendo a la matriz kernel como $\mathbb{K} = \mathbb{X}\mathbb{X}^t$ y \mathbb{D} como la matriz asociada a las distancias, la ecuación 8 puede escribirse como en la ecuación 9.

$$\min \|\mathbb{K}_c - \mathbb{K}^*\| \quad \mathbb{K}_c = \mathbb{C}\mathbb{K}\mathbb{C} \quad (9)$$

La solución de la ecuación 9 es $\mathbb{K} = \sum_i^p \lambda_i v_i v_i^t$, donde v_i y λ_i son los eigenvectores y eigenvalores de \mathbb{K} respectivamente.

2.1.3.1 Kernel lineal

El kernel lineal se encuentra definido en la ecuación 10.

$$K(x, x') = x \cdot x' \quad (10)$$

2.1.3.2 Kernel coseno

El kernel coseno se encuentra definido en la ecuación 11.

$$K(x, x') = \frac{x \cdot x'}{\|x\| \|x'\|} \quad (11)$$

Donde r es un parámetro y d es el grado del polinomio.

2.1.3.3 Kernel gaussiano

El kernel gaussiano se encuentra definido en la ecuación 12.

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (12)$$

Donde γ es un parámetro que controla el comportamiento del kernel. Cuando es muy pequeño el modelo se aproxima al kernel lineal.

2.1.3.4 Kernel sigmoide

El kernel sigmoide se encuentra definido en la ecuación 13.

$$K(x, x') = \tanh(\gamma(x \cdot x') + r) \quad (13)$$

2.1.3.5 Parámetros

En la tabla 2 se encuentran los parámetros que se usaron para cada kernel antes especificado.

Kernel	Grado	γ	r
Lineal	-	-	-
Coseno	-	-	-
Gaussiano	-	$\frac{1}{n}$	-
Sigmoide	-	$\frac{1}{n}$	1

Tabla 2: Parámetros usados para cada kernel. El simbolo – indica que no es necesario el parámetro en el kernel.

2.2. Algoritmos de clasificación

2.2.1. K-means

El algoritmo de K-means es un algoritmo de clasificación no supervisada que agrupa objetos en k conjuntos com base a los valores de las variables dadas. El algoritmo asigna a cada elemento a un conjunto minimizando la suma de las distancias de cada objeto con el centroide del conjunto (ecuación 14).

$$\min_{\{r_{nk}\}\{\mu_k\}} \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (14)$$

2.2.2. Hierarchical Cluster

El algoritmo de Hierarchical clustering es un algoritmo que por medio de una función recursiva une a un par de conjuntos a uno solo usando un enlace por distancias.

Existen diversas maneras de medir una distancia, en este trabajo se utilizo la similitud del coseno (ecuación 15) debido a que presento mejores resultados con los datos utilizados.

$$d(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (15)$$

La función para calcular la distancia entre dos conjuntos es llamada linkage function. En este trabajo se empleo la función complete. La ecuación 16 es la definición de linkage complete.

$$d(C_i, C_j) = \max \{d(x, y) : x \in C_i, y \in C_j\} \quad (16)$$

2.2.3. Self-Organizing-Map

Self-Organizing-Map (SOM) es un algoritmo de clasificación no supervisada. Su objetivo es reducir la dimensionalidad del los datos dados preservando la estructura topológica. El algoritmo

minimiza la ecuación 17.

$$C(\{m_l\}, \{l(i)\}) = \sum_i \sum_k h(||c_k - c_{l(i)}||^2) ||x_i - m_k||^2 \quad (17)$$

Donde m_l es el representante más cercano a x_i , $h()$ es una función decreciente y positiva y c_k son las celdas donde se asociara el dato x_i .

2.3. Base de datos de CONAPO

Se descargo la base de datos de índice de marginación por municipio del periodo 1990 a 2020 disponibles desde el portal de la Consejo Nacional de Población (CONAPO)^{25;26}. Los datos que contienen los archivos se encuentran descritos en la tabla 3.

Abreviación	Descripción
CVE_ENT	Clave de entidad federativa
NOM_ENT	Nombre de entidad federativa
CVE_MUN	Clave del municipio
NOM_MUN	Nombre del municipio
POB_TOT	Población total
ANALF	Porcentaje de población analfabeta de 15 años o más
SBASC	Porcentaje de población de 15 años o más sin educación básica
OVSDE	Porcentaje de ocupantes en viviendas particulares habitadas sin drenaje ni excusado
OVSEE	Porcentaje de ocupantes en viviendas particulares habitadas sin energía eléctrica
OVSAE	Porcentaje de ocupantes en viviendas particulares habitadas sin agua entubada
OVPT	Porcentaje de ocupantes en viviendas particulares habitadas con piso de tierra
VHAC	Porcentaje de viviendas particulares con hacinamiento
PL.5000	Porcentaje de población que vive en localidades menores a 5 000 habitantes
PO2SM	Porcentaje de población ocupada con ingresos de hasta 2 salarios mínimos
IM	Índice de marginación
GM	Grado de marginación

Tabla 3: Diccionario de los indicadores socioeconómicos e índice de marginación por municipio.

Se realizaron diversas modificaciones a los datos obtenidos de CONAPO. Las modificaciones que se realizaron son las siguientes:

- Para el archivo que contiene los datos del año 2015 se eliminó la línea 1555 debido a que no se recabaron datos correspondientes al municipio de Nicolás Ruiz, Chiapas.
- Las celdas que contienen el símbolo - fueron eliminadas. Esto es porque el símbolo fue utilizado para señalar que no se recabó el datos correspondiente.
- Los nombres de las cabeceras de columna fue modificado para que fueran igual a las mostradas en la tabla 3.
- Del archivo que contiene los datos del año 2020 se produjo un archivo auxiliar del tipo csv. La información extraída se encuentra en la hoja llamada *IMM_2020*.

En la figura 1 se muestra la frecuencia relativa de los índices de marginación para los 2469 municipios del año 2015 y 2020.

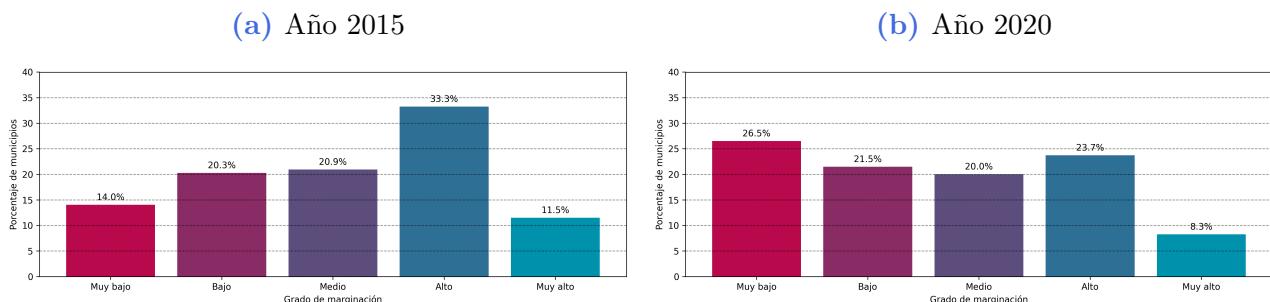


Figura 1: Frecuencia relativa de cada índice de marginación en los años 2015-2020 a nivel municipal.

En la figura 2 se muestran los índices de marginación para cada municipio en los años 2015 y 2020. Los datos para delimitar cada municipio fue obtenido del siguiente blog²⁷. La distribución de los índices de marginación en la República Mexicana marca que se llegan a obtener valores bajos en la zona norte y valores altos en la zona del centro y sur del país.



Figura 2: Grado de marginación a nivel municipio en la República Mexicana.

A partir de las columnas de cada indicador (ANALF, SBASC, OVSDE, OVSEE, OVSAE, OVPT , VHAC , PL.50 y PO2SM) se creó una matriz de embeddings a la cual se normalizó siguiendo la ecuación 3. Esta matriz será la que se le dará como valor de entrada a cada algoritmo presentado en la sección 2.1.

3. Resultados

3.1. Algoritmos de reducción de dimensionalidad

3.1.1. ISOMAP

Se aplicó el algoritmo ISOMAP descrito en la sección 2.1.1 con los datos de índice de marginación para los años 2015 y 2020. En la figura 3 se observa los resultados del algoritmo para el caso bidimensional para 8, 10, 12 y 14 vecinos.

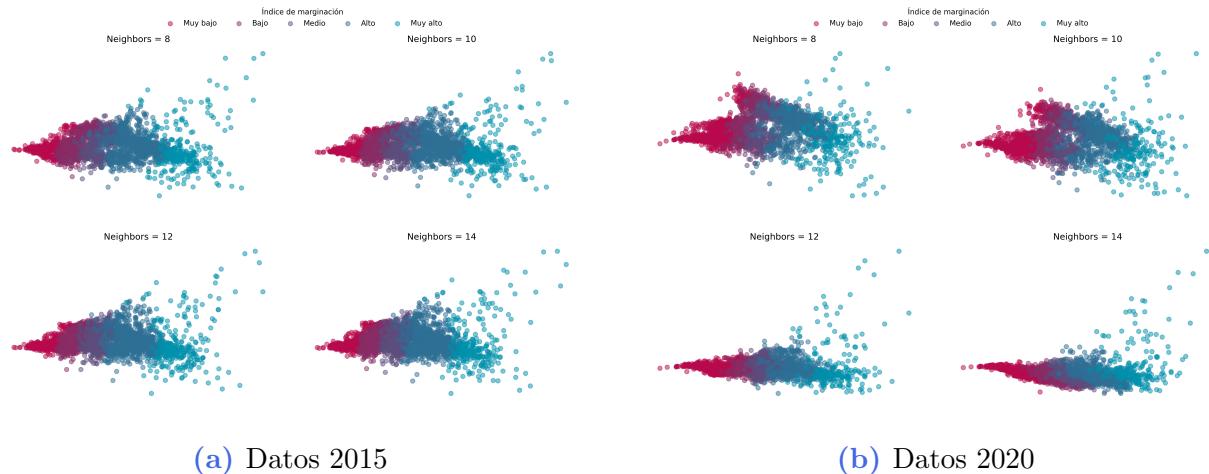


Figura 3: Resultados de aplicar ISOMAP para el caso bidimensional para los datos de índice de marginación de los años 2015 y 2020.

En la tabla 4 se pueden descargar las visualizaciones tridimensionales de los resultados del algoritmo ISOMAP en el caso tridimensional para 8, 10, 12 y 14 vecinos para los años 2015 y 2020.

Vecinos	Años	
	2015	2020
8	Link	Link
10	Link	Link
12	Link	Link
14	Link	Link

Tabla 4: Link de descarga para las visualizaciones tridimensionales de los resultados de ISO-MAP para 8, 10, 12 y 14 vecinos en los años 2015 y 2020.

3.1.2. T-SNE

Se aplicó el algoritmo TSNE descrito en la sección 2.1.2 con los valores de perplejidad de 100, 200, 300 y 400 con los datos de los años 2015 y 2020. En la figura 4 se observan los resultados obtenidos para el caso bidimensional.

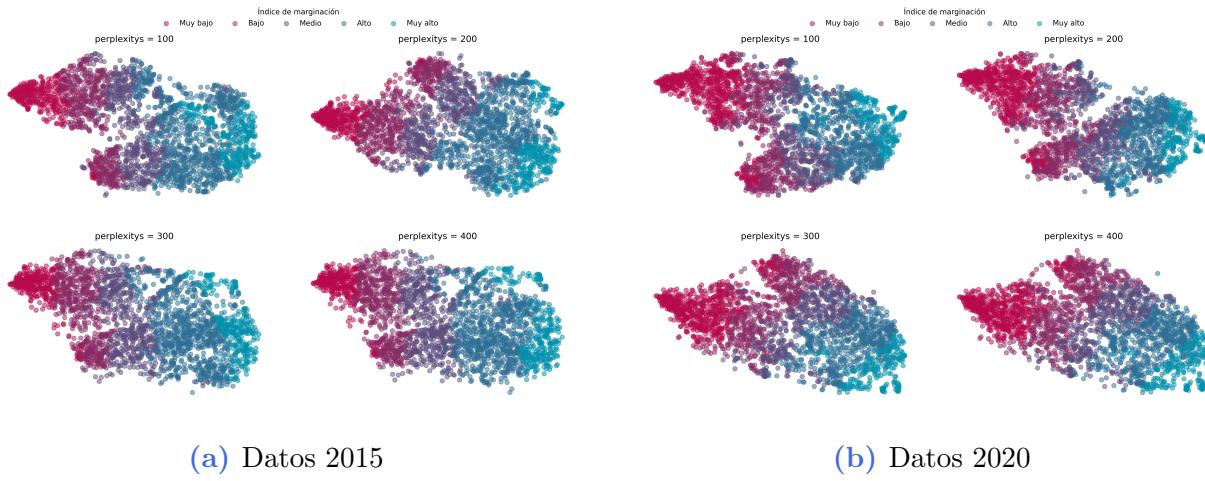


Figura 4: Resultados de aplicar T-SNE para el caso bidimensional para los datos de índice de marginación de los años 2015 y 2020.

En la tabla 5 se pueden descargar las visualizaciones tridimensionales de los resultados del algoritmo ISOMAP en el caso tridimensional.

Perplejidad	Años	
	2015	2020
100	Link	Link
200	Link	Link
300	Link	Link
400	Link	Link

Tabla 5: Link de descarga para las visualizaciones tridimensionales de los resultados de T-SNE para 100, 200, 300 y 400 DE perplejidad en los años 2015 y 2020.

3.1.3. Análisis de componentes principales

Se aplicó el algoritmo de PCA descrito en la sección 2.1.3 con el kernel lineal, polinomial, gaussiano y sigmoide y los parámetros de la tabla 2. En la figura 5 se observan los resultados para el caso bidimensional.

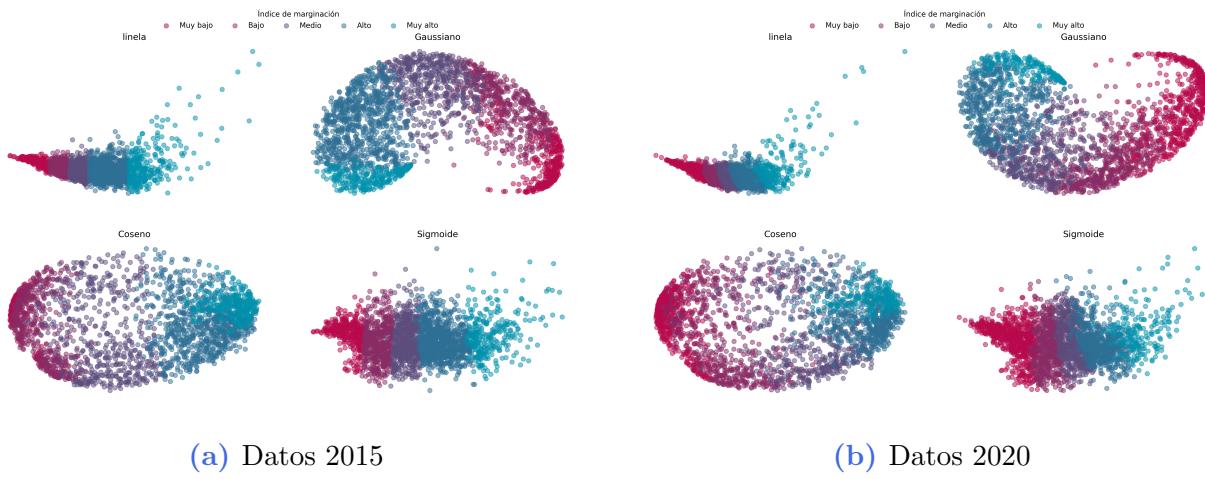


Figura 5: Resultados de aplicar PCA para el caso bidimensional para los datos de índice de marginación de los años 2015 y 2020.

En la tabla 6 se pueden descargar las visualizaciones tridimensionales de los resultados del algoritmo PCA en el caso tridimensional.

Kernel	Años	
	2015	2020
Lineal	Link	Link
Coseno	Link	Link
Gaussiano	Link	Link
Sigmoidal	Link	Link

Tabla 6: Link de descarga para las visualizaciones tridimensionales de los resultados de PCA para el kernel lineal, coseno, gaussiano y sigmoide en los años 2015 y 2020.

3.2. Algoritmos de clasificación

3.2.1. K-means

En la figura 6 se muestran las tablas de confusión resultantes de aplicar el algoritmo de K-means descrito en la sección 2.2.1 usando el inicializador k-means++ y uno aleatorio.

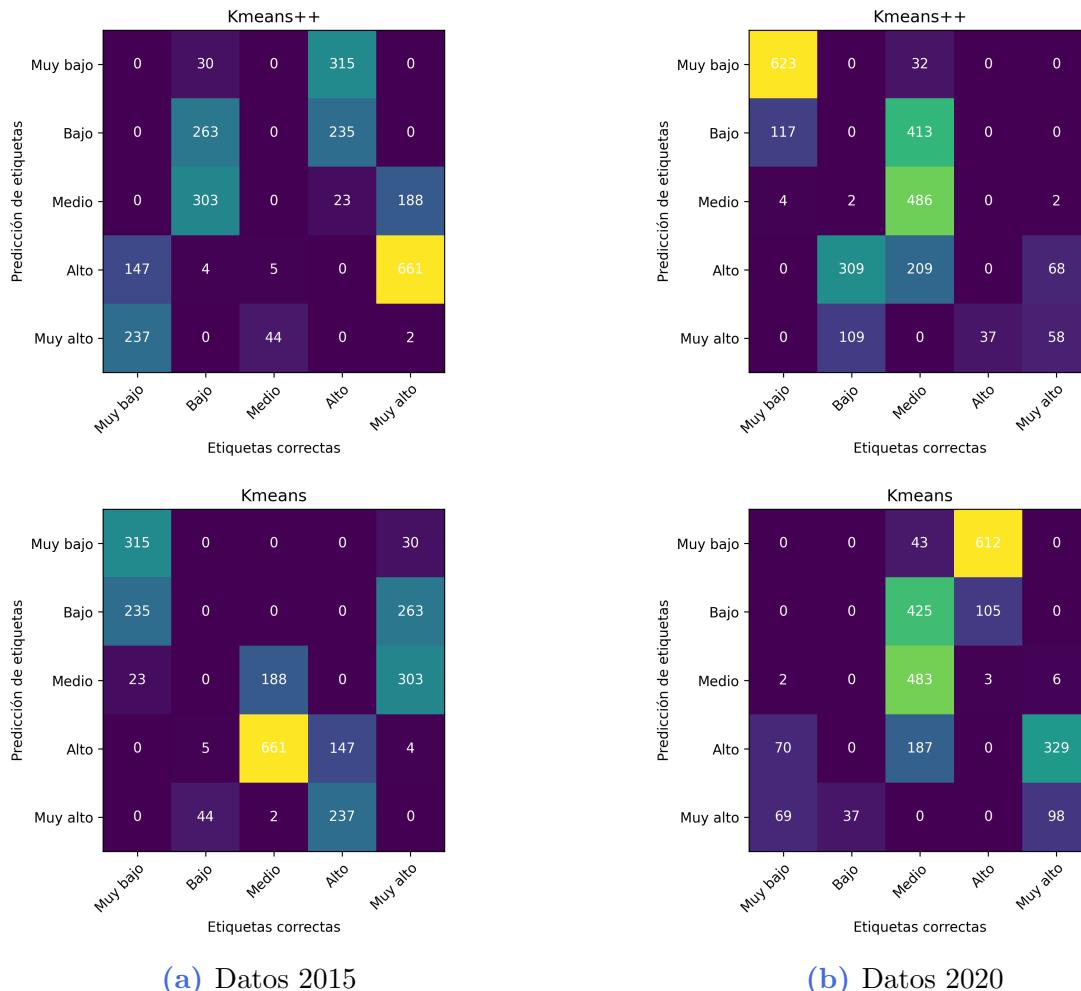
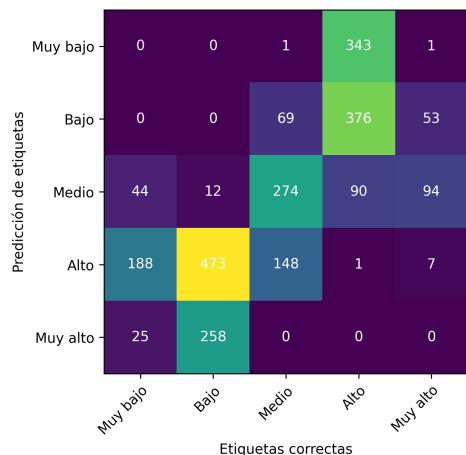


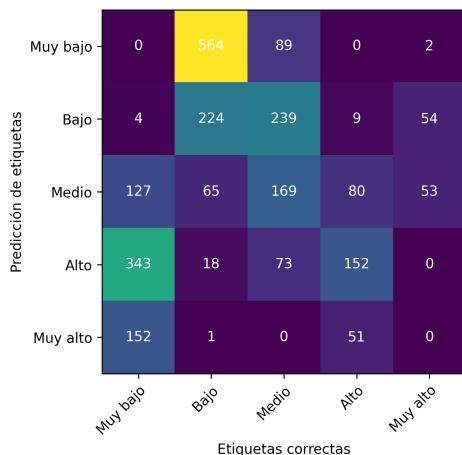
Figura 6: Matrices de confusión resultantes del algoritmo K-means para los datos de índice de marginación de los años 2015 y 2020.

3.2.2. Hierarchical Cluster

En la figura 7 se muestran las tablas de confusión obtenidas con el algoritmo de Hierarchical cluster usando un linkage de tipo complete con una distancia coseno para los años 2015 y 2020.



(a) Año 2015

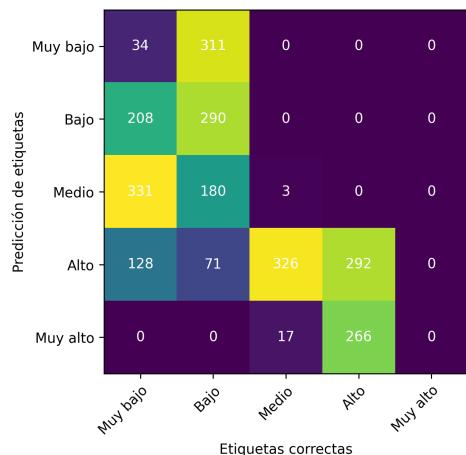


(b) Año 2020

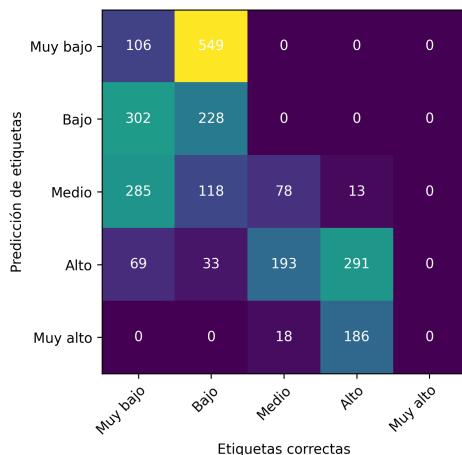
Figura 7: Matrices de confusión resultantes del algoritmo Hierarchical cluster para los años 2015 y 2020.

3.2.3. Self-Organizing-Map

En la figura 8 se muestra la tabla de confusion obtenida a partir de los resultados obtenidos del algoritmo SOM.



(a) Año 2015



(b) Año 2020

Figura 8: Matrices de confusion resultantes para el algoritmo de SOM para los años 2015 y 2020.

4. Conclusiones

4.1. Algoritmos de reducción de dimensionalidad

Realizando una comparación visual con las figuras 3, 4 y 5 se observan las siguientes características:

- Existe una variación en la forma de los conjuntos formados por el algoritmo ISOMAP. Sin embargo, conforme aumenta el número de vecinos, los conjuntos forman estructuras más compactas.
- Los resultados obtenidos por el algoritmo T-SNE presentan diferentes estructuras de los conjuntos para los años 2015 y 2020. Esto puede ser debido a la naturaleza aleatoria del algoritmo. Al aumento en la perplejidad del algoritmo, las estructuras llegan a un equilibrio. Por lo que si se quisiera obtener una reducción de dimensionalidad con una perplejidad mayor, los vectores de posición de cada dato no variaría en mucho con respecto a los presentados en este trabajo.
- A diferencia de los algoritmos de ISOMAP y T-SNE, el algoritmo de PCA obtiene muy buenos resultados al reducir la dimensionalidad de los datos. Esto es debido a que con los kernels utilizados, en todos se puede diferenciar el inicio y el fin de cada conjunto. En especial para el kernel lineal y sigmoide, los cuales pueden separar a cada conjunto por medio de rectas. El mejor de los resultados es el obtenido por el kernel lineal, ya que con solo tomar la primera componente se puede realizar una diferencia en cada conjunto. Esto no es de sorprenderse, debido a que la teoría para formar cada conjunto se basa en este algoritmo.

Realizando una comparación con las animaciones contenidas en las tablas 4, 5 y 6 se observan las siguientes características:

- Los resultados mostrados por ISOMAP no varian tanto de su representación bidimensional. Esto es debido a que la mayoría de los datos están contenidos en un plano independientemente del número de vecinos.
- Para el algoritmo T-SNE se ve beneficiado de tener estructuras más separadas para valores menores de perplejidad en el caso tridimensional, esto es debido a que se logra visualizar el espacio donde se aglomera cada conjunto de datos.
- En el caso del algoritmo de PCA, el kernel gaussiano y coseno se favorecen de una dimensión extra. Esto es debido a que los conjuntos formados son más diferenciables. Para el kernel sigmoide su representación bidimensional es mejor debido a que es más sencillo de visualizar cada conjunto, propiedad que se ve diluida en su representación tridimensional. En cambio, para el kernel lineal la mayoría de los datos se sitúan sobre un plano, por lo que es innecesaria su representación tridimensional.

Con las características antes mencionadas, se obtiene que los mejores resultados son los del algoritmo PCA con kernel lineal y sigmoide debido a que sus visualizaciones diferencian a cada conjunto con únicamente una componente en los dos períodos analizados. El uso del algoritmo de T-SNE para el caso bidimensional con una perplejidad igual o mayor a 300 sería una buena alternativa pero perdería la reproducibilidad de los resultados en un mismo periodo.

4.2. Algoritmos de clasificación

Para los algoritmos de clasificación presentados en este trabajo se encontro que ninguno logra obtener un resultado bueno. Esto se refleja en las tablas de confusión mostradas en las figuras 6, 7 y 8. El algoritmo que obtiene un número mayor de aciertos es el SOM. Con estos resultados se observa que el uso de algoritmos de clasificación no es la mejor opción debido a su baja presición de resultados.

5. Código

Los códigos implementados en este documento se encuentra en el siguiente repositorio [GitHub](#).

6. Referencias

- [1] Consejo Nacional de Población (CONAPO). Índices de marginación 2020; 2021. Available from: https://www.gob.mx/cms/uploads/attachment/file/685354/Nota_tecnica_IMEyM_2020.pdf.
- [2] Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Los mapas de Pobreza en México. Anexo técnico metodológico; 2007. Available from: https://www.coneval.org.mx/rw/resource/coneval/med_pobreza/1024.pdf.
- [3] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2012.; 2011. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2011_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2012.pdf.
- [4] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2013.; 2012. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2012_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2013.pdf.
- [5] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2014.; 2013. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2013_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2014.pdf.
- [6] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2015.; 2014. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2014_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2015.pdf.
- [7] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2016.; 2015. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2015_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2016.pdf.
- [8] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2017.; 2016. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2016_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2017.pdf.

- [9] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2018.; 2017. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2017_SEDESOL_Declaratoria_zonas_atencion_prioritaria_2018.pdf.
- [10] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2019.; 2018. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2018_BIENESTAR_Declaratoria_zonas_atencion_prioritaria_2019.pdf.
- [11] Diario Oficial de la Federación (DOF). Decreto por el que se emite la Declaratoria de Zonas de Atención Prioritaria para el año 2020.; 2019. Available from: https://portalsocial.guanajuato.gob.mx/sites/default/files/documentos/2019_BIENESTAR_Declaratoria_zonas_atencion_prioritarias_2020_20191211.pdf.
- [12] Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Evaluación de impacto de la Estrategia 100x100; 2013. Available from: https://www.coneval.org.mx/Informes/Evaluacion/Impacto/Evaluacion_de_impacto_de_la_Estrategia_100x100.pdf.
- [13] Diario Oficial de la Federación (DOF). Reglas de Operación del Programa de Vivienda Social.; 2020. Available from: https://www.dof.gob.mx/nota_detalle.php?codigo=5585435&fecha=04/02/2020.
- [14] Subsecretaría de Desarrollo Social y Humano (SEDESOL). Análisis de funcionamiento y operación del programa de apoyo alimentario en zonas de atención prioritaria; 2008. Available from: https://www.coneval.org.mx/Informes/Evaluacion/Complementarias/Complementarias_2008/SEDESOL/compl_08_sedesol_PAAZAP.pdf.
- [15] Diario Oficial de la Federación (DOF). Acuerdo por el que se emiten las Reglas de Operación del Programa para el Desarrollo de Zonas Prioritarias, para el ejercicio fiscal 2015; 2014. Available from: https://www.gob.mx/cms/uploads/attachment/file/26103/rop_pdzp.pdf.
- [16] Consejo Nacional de Población (CONAPO). Índices de marginación; 2021. Available from: https://www.gob.mx/cms/uploads/attachment/file/685354/Nota_tecnica_IMEyM_2020.pdf.
- [17] Trapero XBP. Problemas de la medición del bienestar y conceptos afines: (una aplicación al caso español). Presidencia del Gobierno, Instituto Nacional de Estadística; 1977.
- [18] Somarriba N, Pena B. Synthetic Indicators of Quality of Life in Europe. Social Indicators Research. 2008 nov;94(1):115–133. Available from: <https://doi.org/10.1007%2Fs11205-008-9356-y>.
- [19] Zarzosa P. The social welfare in Spain before the crisis: Territorial and chronological analysis. International Journal of Advances in Management and Economics. 2012;1(4):165–171.
- [20] Espina Z, del Pilar M. Aproximación a la medición del bienestar social. Idoneidad del indicador sintético "Distancia-P2". Consejo Superior de Investigaciones Científicas. 1996;
- [21] Nayak P, Mishra SK. Efficiency of Pena's P2 Distance in Construction of Human Development Indices. SSRN Journal Electronic Journal. 2012;Available from: <https://doi.org/10.2139%2Fssrn.2066567>.

- [22] Espina PZ, Arechavala NS. An Assessment of Social Welfare in Spain: Territorial Analysis Using a Synthetic Welfare Indicator. *Soc Indic Res.* 2012 feb;111(1):1–23. Available from: <https://doi.org/10.1007%2Fs11205-012-0005-0>.
- [23] Somorriba Arechavala N, Zarzosa Espina P, Pena Trapero B. La calidad de vida en la unión europea. Un análisis temporal por medio de indicadores sintéticos; 2013. Available from: <https://old.reunionesdeestudiosregionales.org/Oviedo2013/htdocs/pdf/p851.pdf>.
- [24] Dalenius T, Hodges JL. Minimum Variance Stratification. *Journal of the American Statistical Association.* 1959;54(285):88–101. Available from: <http://www.jstor.org/stable/2282141>.
- [25] Consejo Nacional de Población (CONAPO). Índice de marginación por municipio, 1990-2015;. Available from: http://www.conapo.gob.mx/work/models/CONAPO/Marginacion/Datos_Abiertos/Municipio/Base_Indice_de_marginacion_municipal_90-15.csv.
- [26] Consejo Nacional de Población (CONAPO). Índice de marginación por municipio 2020;. Available from: http://www.conapo.gob.mx/work/models/CONAPO/Marginacion/Datos_Abiertos/Municipio/IMM_2020.xls.
- [27] Santos Ochoa JJ. Delimitaciones de los municipios de México; 2021. Available from: http://jjsantoso.com/blog/datos/mapa_mexico.zip.