

INSTITUTO POLITÉCNICO NACIONAL  
UNIDAD PROFESIONAL INTERDISCIPLINARIA DE  
INGENIERÍA  
CAMPUS ZACATECAS



**RECONOCIMIENTO DE PATRONES**

---

**PRÁCTICA 1: CLASIFICACIÓN SUPERVISADA**

---

**Autor:** Manuel Arguijo Landeros

**Docente:** Roberto Oswaldo Cruz Leija

# **Introducción**

Esta práctica tiene como fin introducirnos al estudio de la ciencia informática reconocimiento de patrones la cual consiste en la extracción de información en base a características específicas de un conjunto de objetos. El sistema de reconocimiento debe asignar a cada objeto su categoría o clase (conjunto de entidades que comparten alguna característica que las diferencia del resto). Para poder reconocer los patrones se siguen los siguientes procesos:

1. adquisición de datos.
2. extracción de características.
3. toma de decisiones.

El punto esencial del reconocimiento de patrones es la clasificación. Para la siguiente práctica haremos uso de dos métodos de clasificación supervisada, la cual consiste en un previo entrenamiento con una base de datos ya clasificada, a partir de esta tomará diversas acciones para clasificar las posteriores entradas. En nuestro caso utilizaremos 2 algoritmos para realizar el trabajo; Mínima distancia y vecinos cercanos (knn). Posteriormente mediante el uso de una matriz de confusión (herramienta que permite la visualización del desempeño de un algoritmo) evaluaremos nuestra selección de características y determinaremos cuál o cuáles son más eficaces.

# **Objetivo**

Desarrollar una base de datos que conste de una serie de características y una clase de un determinado objeto para luego mediante el uso de un programa llevar a cabo el proceso de clasificación supervisada de dichos miembros de la base de datos.

# **Materiales**

Computadora y un entorno de desarrollo integrado (Netbeans) además de diversos materiales necesarios para la extracción de características del objeto a clasificar, En este caso:Cuchillo, regla, cinta flexible, báscula de cocina.

## Desarrollo

Para el desarrollo de esta práctica decidí usar como objeto de muestreo la fruta tuna ya que es accesible el adquirir ejemplares de 3 variedades distintas; Blancas, amarillas y rojas. Además de que cada variedad presenta ligeras diferencias de sus características unas de otras, así como diferencias significativas. La extracción de características consistirá en las medidas físicas de cada ejemplar: Altura, Perímetro de su centro y peso, tanto para el fruto con cáscara como sin esta. La base de datos contará con 90 registros, 30 para cada uno de los tipos de tunas.



Figure 1: Muestra del Dataset

Con cáscara			Sin cáscara			
Altura(cm)	Perímetro(cm)	Peso(g)	Altura(cm)	Perímetro(cm)	Peso(g)	Clase
8.3	18.9	151	7.2	16.7	96	Blanca
9.3	20.8	216	7.3	16.7	105	Amarilla
5.4	13.4	51	3.9	11.3	27	Roja

Una vez terminado el proceso de muestreo se deberá de guardar la base de datos en un archivo de texto especificado en el programa de clasificación, luego se deberá de analizar los resultados a través de la matriz de confusión. Así se probarán todas las características; altura, perímetro y peso con y sin cáscara con el fin de encontrar la o las características que mayor porcentaje de éxito resulte.

# Resultados

Mínima distancia	{1,1,1,1,1,1}	KNN
22.0, 8.0, 0.0,		29.0, 1.0, 0.0,
7.0, 23.0, 0.0,		4.0, 26.0, 0.0,
0.0, 0.0, 30.0,		0.0, 0.0, 30.0,
Promedio :83.33 %		Promedio :94.44 %

Podemos ver a través de la matriz de confusión que utilizando todas nuestras características, los algoritmos nos dan una eficacia de 83.33 para mínima distancia y 94.44 para vecinos cercanos. La confusión se da en ambos casos por las tunas amarillas y blancas ya que a simple vista podemos ver que tienen características físicas muy similares, en cambio las tunas rojas son considerablemente distintas, lo cual se ve reflejado en los resultados.

Mínima Distancia						
Con Cáscara			Sin Cáscara			
Altura	Circunferencia	Peso	Altura	Circunferencia	Peso	Clase
63.3%	70%	53.33%	56.66%	90%	76.67%	Blanca
53.3%	70%	53.33%	56.66%	86.67%	83.33%	Amarilla
100%	100%	100%	100%	100%	100%	Roja
72.22%	80%	68.89%	71.11%	92.22%	86.67%	<b>Total</b>

KNN						
Con Cáscara			Sin Cáscara			
Altura	Circunferencia	Peso	Altura	Circunferencia	Peso	Clase
80%	86.67%	86.66%	90%	90%	83.33%	Blanca
50%	60%	66.66%	36.67%	83.33%	80%	Amarilla
100%	100%	100%	100%	100%	100%	Roja
76.67%	82.2%	84.44%	75.56%	91.11%	87.78%	<b>Total</b>

Figure 2: Porcentajes de eficacia por caracteristica

La tabla muestra los porcentajes de eficacia de cada una de las características analizadas por separado. Podemos ver que la circunferencia del fruto sin cáscara es la que mejor resultados nos ofrece, a su vez los que peor resultados da son; el peso del fruto con cáscara en el algoritmo de mínima distancia y la altura sin cáscara para el de vecinos cercanos.

Nuestro programa cuenta con un arreglo binario que nos permite seleccionar las características deseadas, en este caso se tiene un total de 6. Por tanto existen 63 posiciones posibles en los que al menos 1 característica sea tomada en cuenta. Al analizar cada una de ellas determiné el conjunto de características que mayor porcentaje de acierto ofreciera para cada algoritmo:

$\{1,0,0,0,1,0\}$	$\{0,1,0,0,1,0\}$
<b>Mínima distancia</b>	<b>KNN</b>
27.0, 3.0, 0.0,	29.0, 1.0, 0.0,
4.0, 26.0, 0.0,	3.0, 27.0, 0.0,
0.0, 0.0, 30.0,	0.0, 0.0, 30.0,
<b>Promedio : 92.22 %</b>	<b>Promedio : 95.56 %</b>

El conjunto de características que mejores resultados nos da son: la altura con cáscara y el perímetro sin cáscara(para mínima distancia) y el perímetro sin cáscara y perímetro con cáscara(para KNN). En las siguientes gráficas podemos apreciar la distribución de las características con mejores resultados para cada caso:

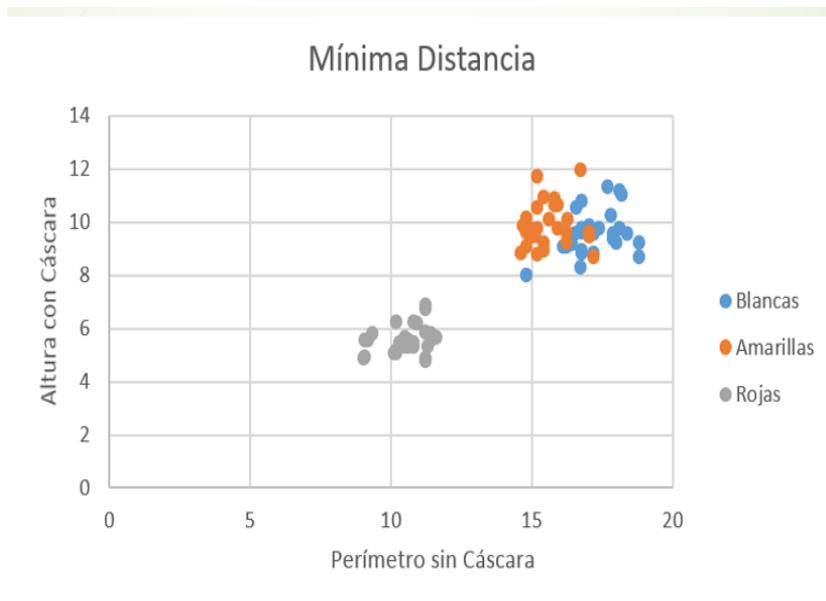


Figure 3: Gráfica de distribución de características

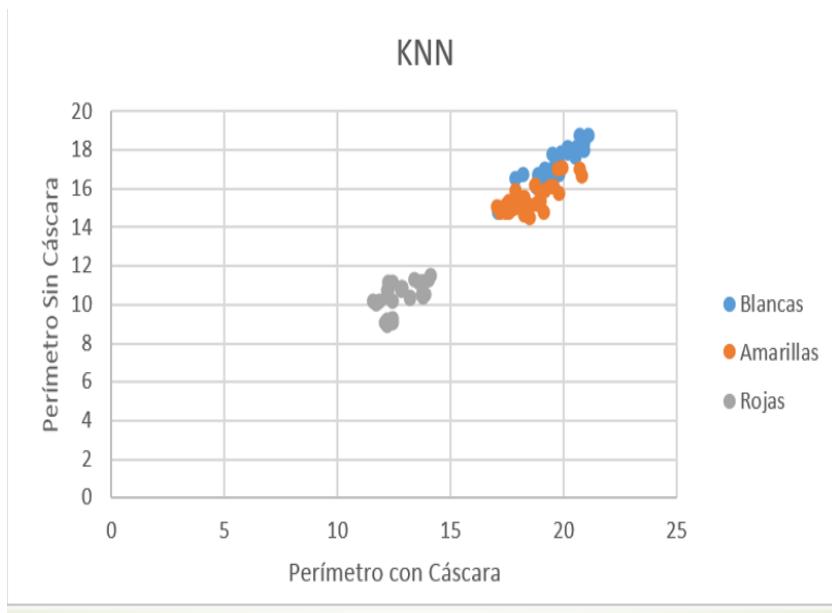


Figure 4: Gráfica de distribución de características

## Conclusión

Se cumplió con el objetivo de la práctica ya que la base de datos se logró desarrollar adecuadamente con un número suficiente de ejemplares muestrados. Tampoco hubo complicaciones con la ejecución del programa ni de los 2 algoritmos utilizados, además los resultados corroboraron una previa hipótesis acerca de las diferencias de densidad de las cáscaras de dos de los ejemplares.